US 20190177777A1

(19) United States

(12) Patent Application Publication

CHEE

(10) Pub. No.: US 2019/0177777 A1

(43) Pub. Date: Jun. 13, 2019

(54) **SPATIALLY ENCODED BIOLOGICAL ASSAYS**

(71) Applicant: **Prognosys Biosciences, Inc.**, San Diego, CA (US)

(72) Inventor: **Mark S. CHEE**, San Diego, CA (US)

(73) Assignee: **Prognosys Biosciences, Inc.**, San Diego, CA (US)

(21) Appl. No.: **16/276,235**

(22) Filed: **Feb. 14, 2019**

Related U.S. Application Data

(63) Continuation of application No. 15/187,661, filed on Jun. 20, 2016, which is a continuation of application No. 13/080,616, filed on Apr. 5, 2011, now Pat. No. 9,371,598.

(60) Provisional application No. 61/321,124, filed on Apr. 5, 2010.

Publication Classification

(51) **Int. Cl.**

C12Q 1/6837	(2006.01)
C12Q 1/6869	(2006.01)
C12Q 1/6841	(2006.01)
C12Q 1/6834	(2006.01)
C12Q 1/6809	(2006.01)
G01N 33/68	(2006.01)
C40B 60/04	(2006.01)
C40B 30/04	(2006.01)
G01N 33/53	(2006.01)
C12Q 1/6874	(2006.01)
C12Q 1/6804	(2006.01)

(52) **U.S. Cl.**

CPC ***C12Q 1/6837*** (2013.01); ***C12Q 1/6869*** (2013.01); ***C12Q 1/6841*** (2013.01); ***C12Q 1/6834*** (2013.01); ***C12Q 1/6809*** (2013.01); *G01N 2458/10* (2013.01); ***C40B 60/04*** (2013.01); ***C40B 30/04*** (2013.01); ***G01N 33/5308*** (2013.01); ***C12Q 1/6874*** (2013.01); ***C12Q 1/6804*** (2013.01); ***G01N 33/6845*** (2013.01)

(57) **ABSTRACT**

The present invention provides assays and assay systems for use in spatially encoded biological assays. The invention provides an assay system comprising an assay capable of high levels of multiplexing where reagents are provided to a biological sample in defined spatial patterns; instrumentation capable of controlled delivery of reagents according to the spatial patterns; and a decoding scheme providing a readout that is digital in nature.

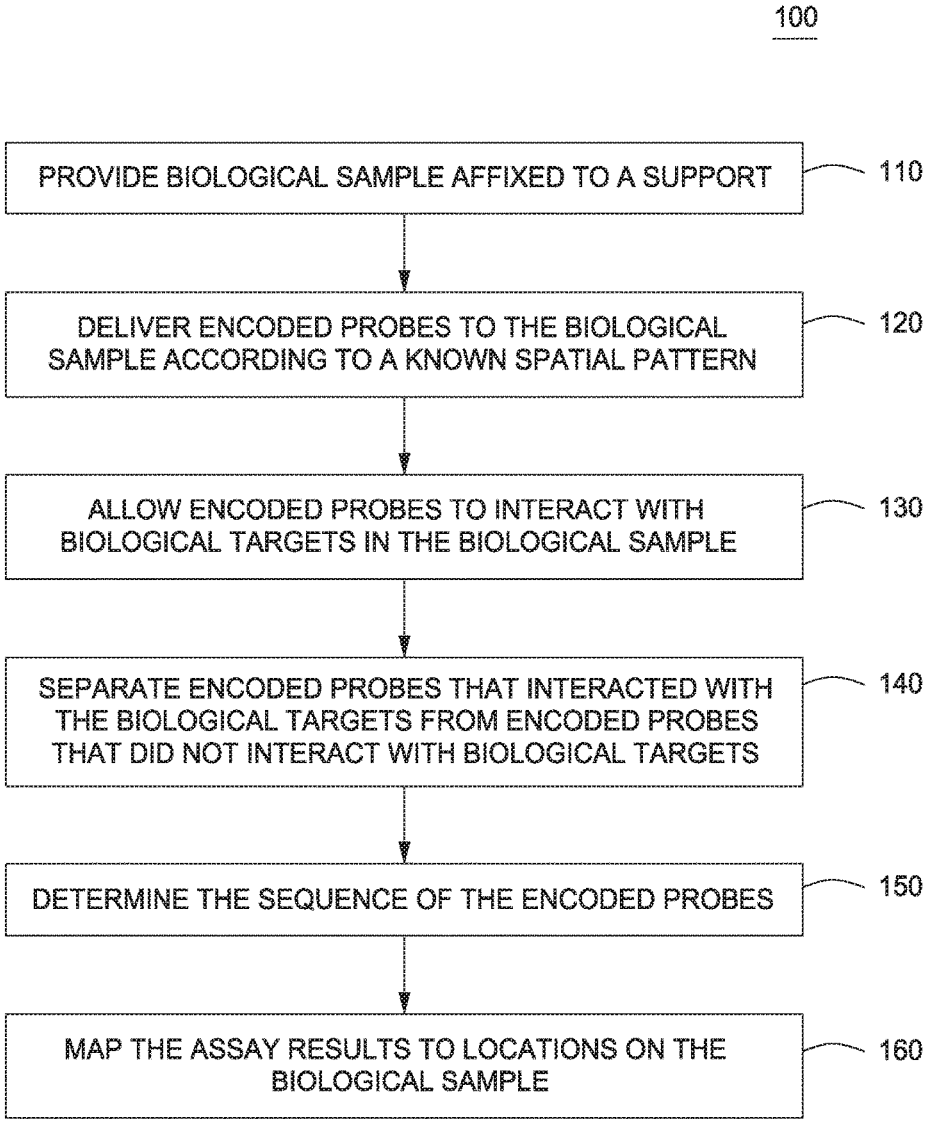
100
PROVIDE BIOLOGICAL SAMPLE AFFIXED TO A SUPPORT
110
DELIVER ENCODED PROBES TO THE BIOLOGICAL SAMPLE ACCORDING TO A KNOWN SPATIAL PATTERN
120
ALLOW ENCODED PROBES TO INTERACT WITH BIOLOGICAL TARGETS IN THE BIOLOGICAL SAMPLE
130
SEPARATE ENCODED PROBES THAT INTERACTED WITH THE BIOLOGICAL TARGETS FROM ENCODED PROBES THAT DID NOT INTERACT WITH BIOLOGICAL TARGETS
140
DETERMINE THE SEQUENCE OF THE ENCODED PROBES
150
MAP THE ASSAY RESULTS TO LOCATIONS ON THE BIOLOGICAL SAMPLE
160

FIG. 1

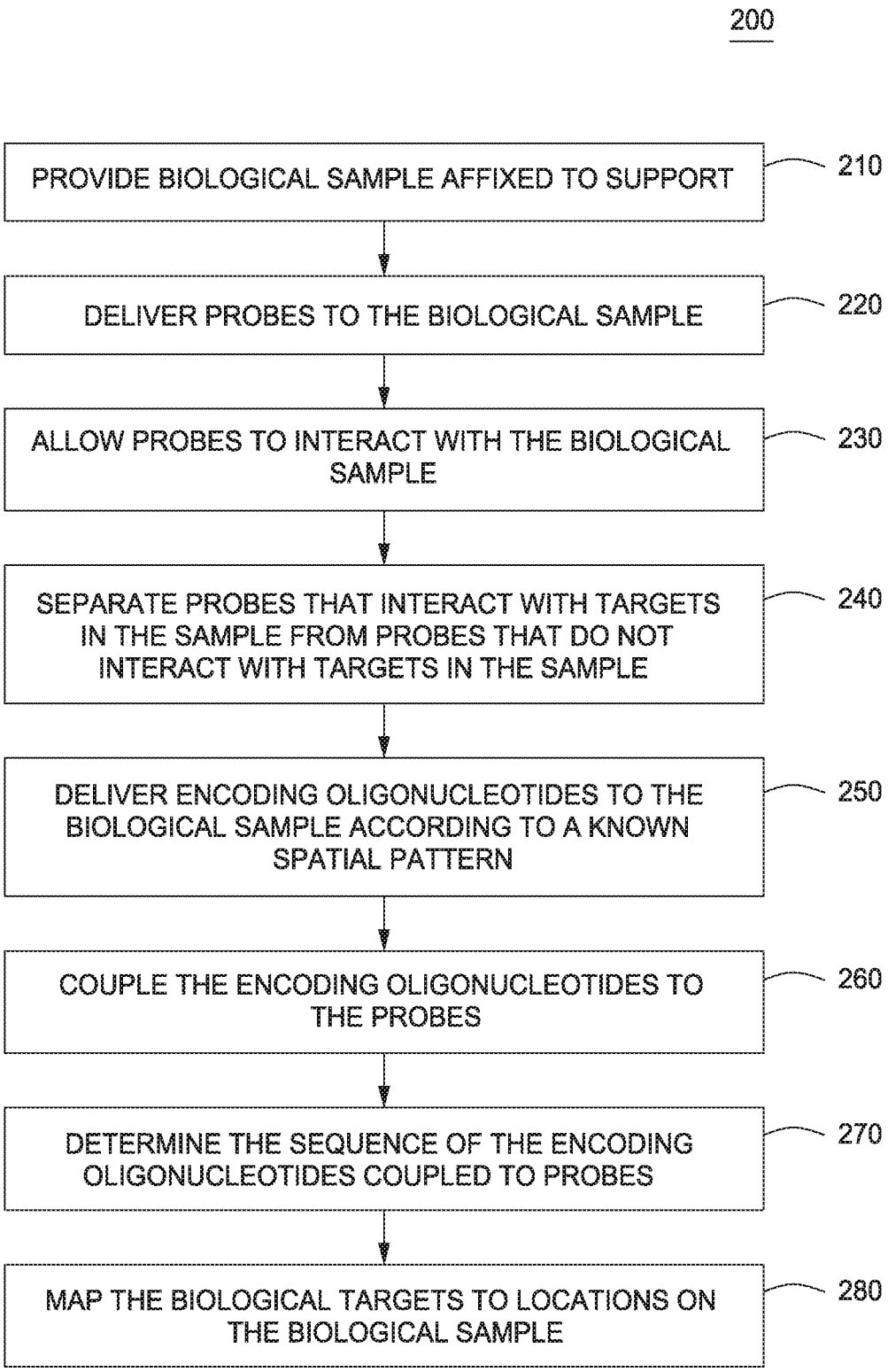
200
PROVIDE BIOLOGICAL SAMPLE AFFIXED TO SUPPORT
210
DELIVER PROBES TO THE BIOLOGICAL SAMPLE
220
ALLOW PROBES TO INTERACT WITH THE BIOLOGICAL SAMPLE
230
SEPARATE PROBES THAT INTERACT WITH TARGETS IN THE SAMPLE FROM PROBES THAT DO NOT INTERACT WITH TARGETS IN THE SAMPLE
240
DELIVER ENCODING OLIGONUCLEOTIDES TO THE BIOLOGICAL SAMPLE ACCORDING TO A KNOWN SPATIAL PATTERN
250
COUPLE THE ENCODING OLIGONUCLEOTIDES TO THE PROBES
260
DETERMINE THE SEQUENCE OF THE ENCODING OLIGONUCLEOTIDES COUPLED TO PROBES
270
MAP THE BIOLOGICAL TARGETS TO LOCATIONS ON THE BIOLOGICAL SAMPLE
280

FIG. 2

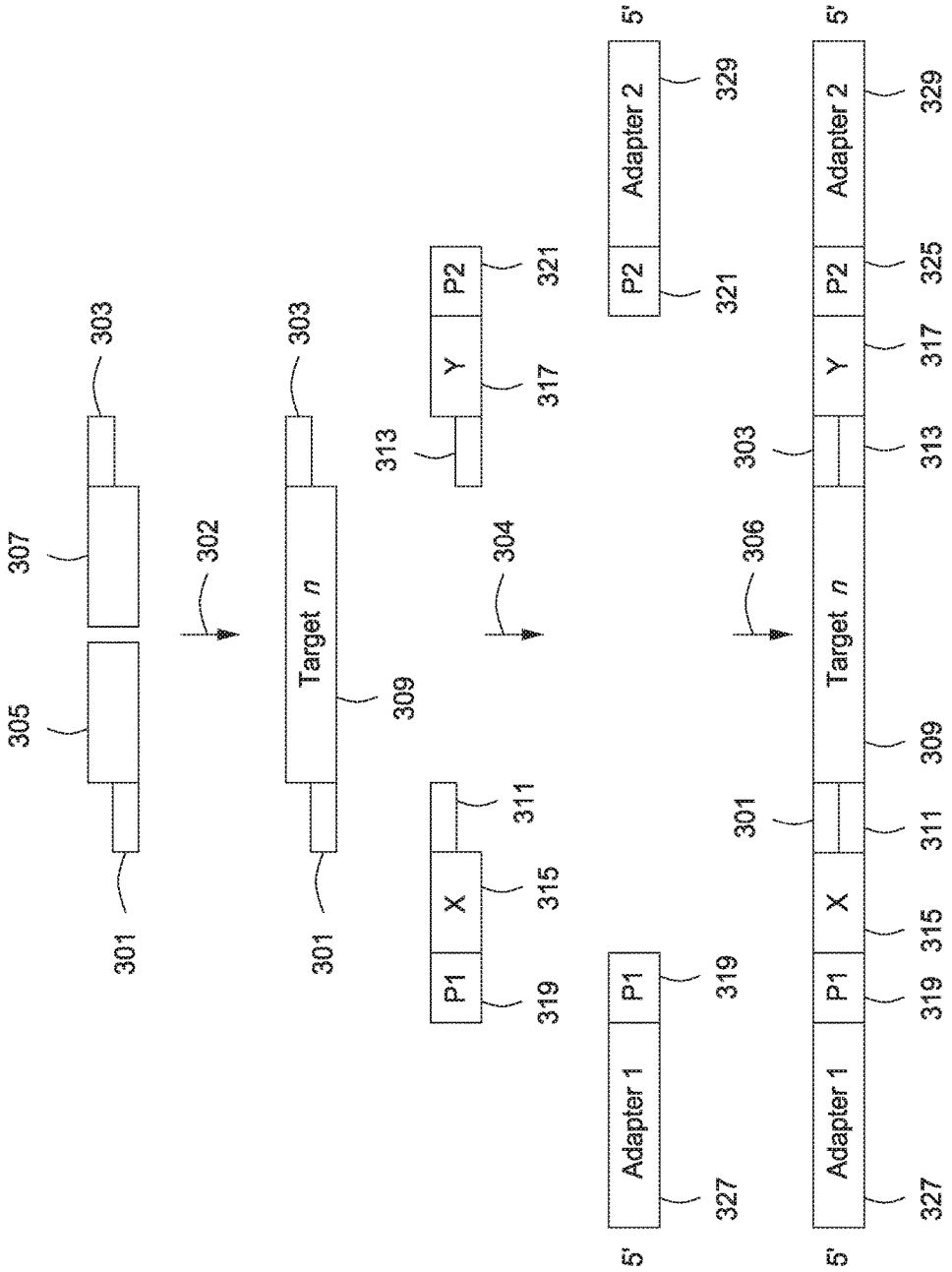
301
302
303
305
307
Target n
309
304
311
313
315
317
319
321
P1
X
Y
P2
Adapter 1
Adapter 2
5'
327
329
306
325

FIG. 3

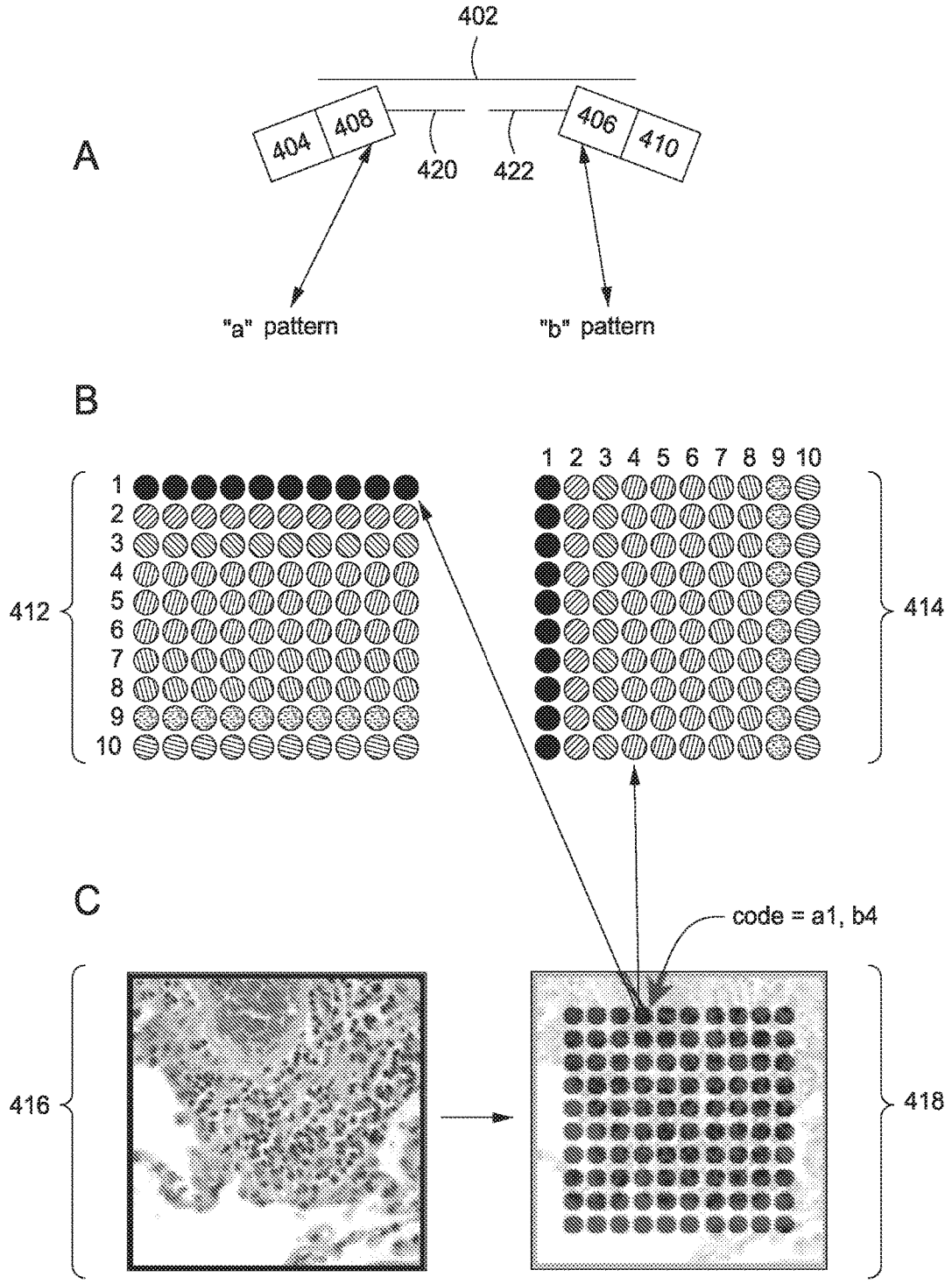
402
A
404
408
406
410
420
422
"a" pattern
"b" pattern
B
1 2 3 4 5 6 7 8 9 10
412
414
C
code = a1, b4
416
418

FIG. 4

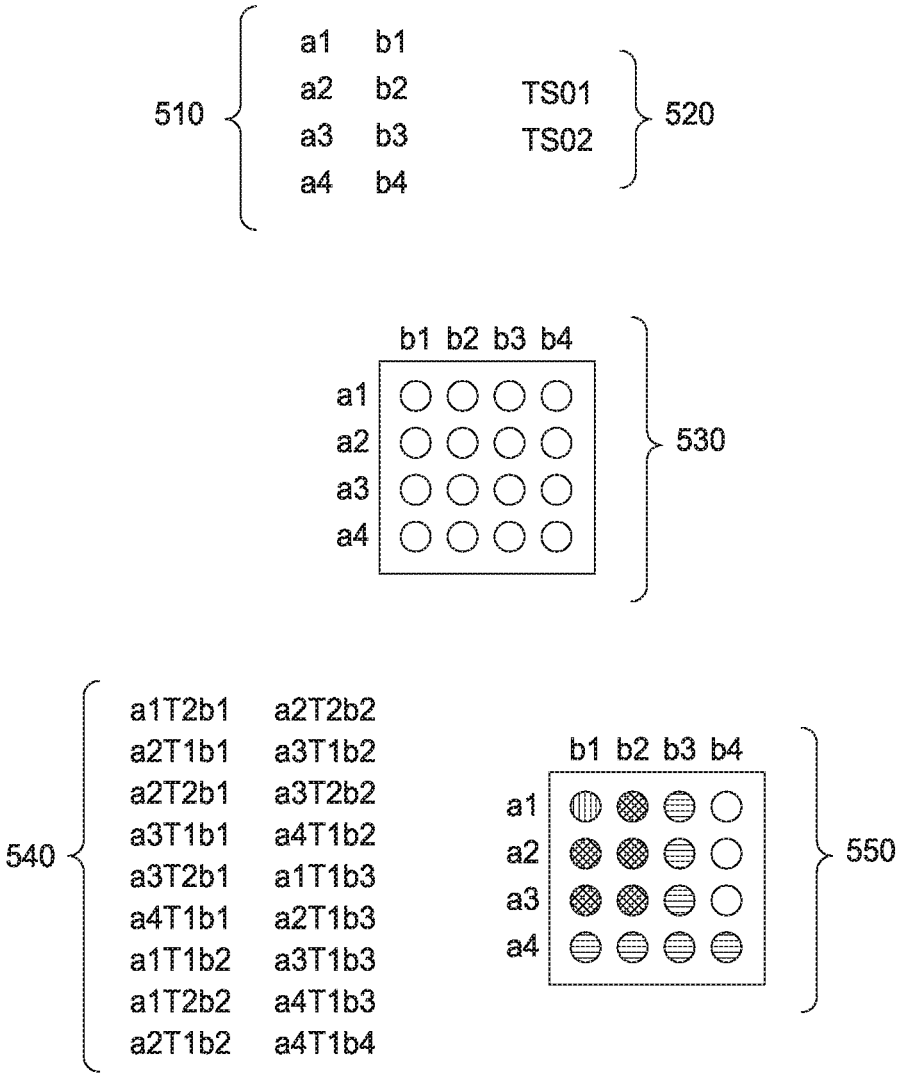
510
a1 b1
a2 b2
a3 b3
a4 b4
TS01
TS02
520
b1 b2 b3 b4
a1
a2
a3
a4
530
540
a1T2b1 a2T2b2
a2T1b1 a3T1b2
a2T2b1 a3T2b2
a3T1b1 a4T1b2
a3T2b1 a1T1b3
a4T1b1 a2T1b3
a1T1b2 a3T1b3
a1T2b2 a4T1b3
a2T1b2 a4T1b4
b1 b2 b3 b4
a1
a2
a3
a4
550

FIG. 5

SPATIALLY ENCODED BIOLOGICAL ASSAYS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 61/321,124, filed Apr. 5, 2010 and is incorporated herein by reference.

FIELD OF THE INVENTION

[0002] This invention relates to assays of biological molecules, and more particularly to assays for determining spatial distributions of a large number of biological molecules in a solid sample simultaneously.

BACKGROUND OF THE INVENTION

[0003] In the following discussion certain articles and methods will be described for background and introductory purposes. Nothing contained herein is to be construed as an "admission" of prior art. Applicant expressly reserves the right to demonstrate, where appropriate, that the articles and methods referenced herein do not constitute prior art under the applicable statutory provisions.

[0004] Comprehensive gene expression analysis and protein analysis have been useful tools in understanding mechanisms of biology. Use of these tools has allowed the identification of genes and proteins involved in development and in various diseases such as cancer and autoimmune disease. Conventional methods such as in situ hybridization and other multiplexed detection of different transcripts have revealed spatial patterns of gene expression and have helped shed light on the molecular basis of development and disease. Other technologies that have enabled the quantitative analysis of many RNA sequences per sample include microarrays (see Shi, et al., Nature Biotechnology, 24(9): 1151-61 (2006); and Slonim and Yanai, Plos Computational Biology, 5(10):e1000543 (2009)); serial analysis of gene expression (SAGE) (see Velculescu, et al, Science, 270 (5235):484-87 (1995)), high-throughput implementations of qPCR (see Spurgeon, et al., Plos ONE, 3(2):e1662 (2008)) and in situ PCR (see Nuovo, Genome Res., 4:151-67 (1995)). As useful as these methods are, however, they do not enable simultaneous measurement of the expression of many genes or the presence and/or activity of multiple proteins at many spatial locations in a sample. Laser capture microdissection has permitted the analysis of many genes at a small number of locations, but it is very expensive, laborious, and does not scale well. Certain PCR assays in a 2D format preserve spatial information (see Armani, et al., Lab on a Chip, 9(24): 3526-34 (2009)), but these methods have low spatial resolution because they rely on physical transference of tissue into wells, which also prevents random access to tissue samples and high levels of multiplexing.

[0005] At present, no practical method exists to analyze at high resolution the spatial expression patterns of large numbers of genes, proteins, or other biologically active molecules simultaneously. There is thus a need for reproducible, high-resolution spatial maps of biological molecules in tissues. The present invention addresses this need.

SUMMARY OF THE INVENTION

[0006] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter. Other features, details, utilities, and advantages of the claimed subject matter will be apparent from the following written Detailed Description including those aspects illustrated in the accompanying drawings and defined in the appended claims.

[0007] The invention encompasses assay systems that provide high-resolution spatial maps of biological activity in tissues. The assay system comprises an assay capable of high levels of multiplexing where encoded probes are provided to a biological sample in defined spatial patterns; instrumentation capable of controlled delivery of reagents according to the spatial patterns; and a decoding scheme providing a readout that is digital in nature. In short, the present invention provides the ability to look at many biological targets in many locations, providing the resolution of in situ hybridization with the highly-parallel data analysis of sequencing.

[0008] Thus, in some embodiments, the invention provides an assay system to determine spatial patterns of abundance or activity or both of multiple biological targets at multiple sites in a sample, where the assay system performs the following steps: providing a sample affixed to a support; delivering encoded probes for the multiple biological targets to the multiple sites in the sample in a known spatial pattern, where each encoded probe comprises a probe region that may interact with the biological targets and a coding tag that identifies a location of the site to which the encoded probe was delivered; allowing the encoded probes to interact with the biological targets; separating encoded probes that interact with the biological targets from encoded probes that do not interact with the biological targets; determining all or a portion of a sequence of the encoded probes, and associating the abundance or activity or both of the multiple biological targets to the locations of the sites in the sample.

[0009] In particular aspects of the invention the biological targets comprise nucleic acids and the encoded probes are oligonucleotides, and in some aspects, there are two encoded probes for each of the multiple nucleic acid targets. In some aspects, the multiple biological targets comprise proteins, the probe regions of the encoding probes are proteins and the coding tags comprise oligonucleotides. In some aspects the multiple biological targets comprise enzymes. In some aspects the probe regions of the encoded probes comprise antibodies, aptamers or small molecules.

[0010] Some aspects of the assay system further comprise an amplification step between the separating step and the determining step. In some aspects, the determining step is performed by nucleic acid sequencing, and in preferred aspects, the sequencing is high-throughput digital nucleic acid sequencing.

[0011] In some aspects of the invention, the product of the multiple biological targets being assayed and the multiple sites in the sample is greater than 20, in some aspects product of the multiple biological targets being assayed and the multiple sites in the sample is greater than 50, in some aspects the product of the multiple biological targets being assayed and the multiple sites in the sample is greater than

75, 100, 150, 500, 750, 1,000, 5,000, 10,000, 25,000, 50,000, 100,000, 500,000, or 1,000,000 or more. In other aspects, the sequence of at least fifty thousand encoding probes are determined in parallel, in other aspects the sequence of at least one hundred thousand encoding probes are determined in parallel, in some aspects the sequence of at least five hundred thousand encoding probes are determined in parallel, and in some aspects the sequence of at least one million, ten million, one hundred million, one billion, ten billion, one hundred billion or more encoding probes are determined in parallel.

[0012] In some aspects, the known spatial pattern is determined by histological features of the sample. Also in some aspects, software programmed hardware performs at least two steps of the delivering step, the separation step, the determining step and the associating step.

[0013] In some aspects, the probe regions of the encoded probes are proteins and the separating step is accomplished by encoded probes that interact with the biological targets being captured by an affinity capture agent. In some aspects the probe regions of the encoding probes are nucleic acids and the separating step is accomplished by a washing of the sample.

[0014] In other embodiments there is provided an assay system to determine spatial patterns of abundance or activity or both of multiple nucleic acid targets at multiple sites in a sample, where the assay system performs the following steps: providing a sample affixed to a support; delivering oligonucleotide probes for multiple nucleic acid targets to the multiple sites in the sample in a known spatial pattern; allowing the oligonucleotide probes to hybridize with the nucleic acid targets; washing unhybridized encoded oligonucleotide probes from the sample; delivering one or more encoding agents to locations of the multiple sites in the sample according to a known spatial pattern, where the combination of encoding agents delivered to each site is different; coupling the encoding agents and the oligonucleotide probes to form encoded probes; determining all or a portion of a sequence of the encoded probes using high-throughput sequencing, and associating the abundance or activity or both of multiple biological targets to the locations of multiple sites in the sample.

[0015] Other embodiments of the invention provide an assay system to determine spatial patterns of abundance or activity or both of multiple protein targets at multiple sites in a sample, where the assay system performs the following steps: providing a sample affixed to a support; delivering encoded probes for the multiple protein targets to the multiple sites in the sample in a known spatial pattern, where each encoded probe comprises a protein probe region that may interact with the protein targets and a coding tag that identifies a location of the site to which the encoded probe was delivered and the protein probe region of the encoding probe of which the coding tag is part; allowing the encoded probes to interact with the protein targets; separating encoded probes that interact with the protein targets from encoded probes that do not interact with the protein targets; determining all or a portion of a sequence of the encoded probes by high throughput sequencing, and associating the abundance or activity or both of the multiple protein targets to the locations of the multiple sites in the sample.

[0016] Other embodiments provide an assay system to determine spatial patterns of abundance or activity or both of multiple biological targets at multiple sites in a sample, where the assay system performs the following steps: providing a sample affixed to a support; delivering encoded probes for the multiple biological targets to the multiple sites in the sample in a known spatial pattern, where each encoded probe comprises a probe region that may interact with the biological targets and a coding tag that identifies a location of the site to which the encoded probe was delivered and identifies the biological target; allowing the encoded probes to interact with the biological targets; determining all or a portion of a sequence of the encoded probes, and associating the abundance or activity or both of the multiple biological targets to the locations of the sites in the sample.

[0017] The assay system of the invention can utilize various detection mechanisms, based on the molecules to be detected and the reagents needed for such detection system. Exemplary methods that can be used with the assay systems of the invention are described in more detail below.

DESCRIPTION OF THE FIGURES

[0018] FIG. **1** provides a simplified overview of the assay system of the present invention.

[0019] FIG. **2** provides a simplified overview of one embodiment of the assay system of the present invention for detecting nucleic acids.

[0020] FIG. **3** is a representational depiction of one embodiment of the assay overviewed in FIG. **2**.

[0021] FIG. **4** illustrates a general mechanism for one embodiment of a combinatorial encoding scheme of the assay systems of the invention.

[0022] FIG. **5** provides a simplified, specific example of the embodiment of a combinatorial encoding scheme shown in FIG. **4**.

DEFINITIONS

[0023] The terms used herein are intended to have the plain and ordinary meaning as understood by those of ordinary skill in the art. The following definitions are intended to aid the reader in understanding the present invention, but are not intended to vary or otherwise limit the meaning of such terms unless specifically indicated.

[0024] The term "antibody" as used herein is intended to refer to an entire immunoglobulin or antibody or any functional fragment of an immunoglobulin molecule which is capable of specific binding to an antigen (antibodies and antigens are "binding partners" as defined herein). "Antibody" as used herein is meant to include the entire antibody as well as any antibody fragments capable of binding the antigen or antigenic fragment of interest. Examples of such peptides include complete antibody molecules, antibody fragments, such as Fab, F(ab')2, CDRS, VL, VH, and any other portion of an antibody which is capable of specifically binding to an antigen. Antibodies for assays of the invention are immunoreactive or immunospecific for, and therefore specifically and selectively bind to, proteins either detected (i.e., biological targets) or used for detection (i.e., probes) in the assays of the invention.

[0025] The term "binding agent" as used herein refers to any agent that specifically binds to a biological molecule of interest

[0026] "Complementary" or "substantially complementary" refers to the hybridization or base pairing or the formation of a duplex between nucleotides or nucleic acids, such as, for instance, between the two strands of a double-

stranded DNA molecule or between an oligonucleotide primer and a primer binding site on a single-stranded nucleic acid. Complementary nucleotides are, generally, A and T (or A and U), or C and G. Two single-stranded RNA or DNA molecules are said to be substantially complementary when the nucleotides of one strand, optimally aligned and compared and with appropriate nucleotide insertions or deletions, pair with at least about 80% of the other strand, usually at least about 90% to about 95%, and even about 98% to about 100%).

[0027] "Hybridization" refers to the process in which two single-stranded polynucleotides bind non-covalently to form a stable double-stranded polynucleotide, The resulting (usually) double-stranded polynucleotide is a "hybrid" or "duplex," "Hybridization conditions" will typically include salt concentrations of approximately less than 1M, often less than about 500 mM and may be less than about 200 mM. A "hybridization buffer" is a buffered salt solution such as 5% SSPE, or other such buffers known in the art. Hybridization temperatures can be as low as 5° C., but are typically greater than 22° C., and more typically greater than about 30° C., and typically in excess of 37° C. Hybridizations are often performed under stringent conditions, i.e., conditions under which a primer will hybridize to its target subsequence but will not hybridize to the other, non-complementary sequences. Stringent conditions are sequence-dependent and are different in different circumstances. For example, longer fragments may require higher hybridization temperatures for specific hybridization than short fragments. As other factors may affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents, and the extent of base mismatching, the combination of parameters is more important than the absolute measure of any one parameter alone. Generally stringent conditions are selected to be about 5° C. lower than the Tm for the specific sequence at a defined ionic strength and pH. Exemplary stringent conditions include a salt concentration of at least 0.01 M to no more than 1M sodium ion concentration (or other salt) at a pH of about 7.0 to about 8.3 and a temperature of at least 25° C. For example, conditions of 5×SSPE (750 mM NaCl, 50 mM sodium phosphate, 5 mM EDTA at pH 7.4) and a temperature of approximately 30° C. are suitable for allele-specific hybridizations, though a suitable temperature depends on the length and/or GC content of the region hybridized.

[0028] "Ligation" means to form a covalent bond or linkage between the termini of two or more nucleic acids, e.g., oligonucleotides and/or polynucleotides, in a template-driven reaction. The nature of the bond or linkage may vary widely and the ligation may be carried out enzymatically or chemically. As used herein, ligations are usually carried out enzymatically to form a phosphodiester linkage between a 5' carbon terminal nucleotide of one oligonucleotide with a 3' carbon of another nucleotide.

[0029] "Nucleic acid", "oligonucleotide", "oligo" or grammatical equivalents used herein refers generally to at least two nucleotides covalently linked together. A nucleic acid generally will contain phosphodiester bonds, although in some cases nucleic acid analogs may be included that have alternative backbones such as phosphoramidite, phosphorodithioate, or methylphophoroamidite linkages; or peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with bicyclic structures including locked nucleic acids, positive backbones, non-ionic backbones and non-ribose backbones. Modifications of the ribose-phosphate backbone may be done to increase the stability of the molecules; for example, PNA:DNA hybrids can exhibit higher stability in some environments.

[0030] "Primer" means an oligonucleotide, either natural or synthetic, that is capable, upon forming a duplex with a polynucleotide template, of acting as a point of initiation of nucleic acid synthesis and being extended from its 3' end along the template so that an extended duplex is formed. The sequence of nucleotides added during the extension process is determined by the sequence of the template polynucleotide. Primers usually are extended by a DNA polymerase.

[0031] The term "SNP" or "single nucleotide polymorphism" refers to a genetic variation between individuals; e.g., a single nitrogenous base position in the DNA of organisms that is variable. SNPs are found across the genome; much of the genetic variation between individuals is due to variation at SNP loci, and often this genetic variation results in phenotypic variation between individuals. SNPs for use in the present invention and their respective alleles may be derived from any number of sources, such as public databases (U.C. Santa Cruz Human Genome Browser Gateway (http://genome.ucsc.edu/cgi-bin/hgGateway) or the NCBI dbSNP website (http://www.ncbi.nlm.nih.gov/SNP/), or may be experimentally determined as described in U.S. Pat. No. 6,969,589; and US Pub. No. 2006/0188875 entitled "Human Genomic Polymorphisms." Although the use of SNPs is described in some of the embodiments presented herein, it will be understood that other biallelic or multi-allelic genetic markers may also be used. A biallelic genetic marker is one that has two polymorphic forms, or alleles. As mentioned above, for a biallelic genetic marker that is associated with a trait, the allele that is more abundant in the genetic composition of a case group as compared to a control group is termed the "associated allele," and the other allele may be referred to as the "unassociated allele." Thus, for each biallelic polymorphism that is associated with a given trait (e.g., a disease or drug response), there is a corresponding associated allele. Other biallelic polymorphisms that may be used with the methods presented herein include, but are not limited to multinucleotide changes, insertions, deletions, and translocations. It will be further appreciated that references to DNA herein may include genomic DNA, mitochondrial DNA, episomal DNA, and/or derivatives of DNA such as amplicons, RNA transcripts, cDNA, DNA analogs, etc. The polymorphic loci that are screened in an association study may be in a diploid or a haploid state and, ideally, would be from sites across the genome.

[0032] The term "selectively binds", "selective binding" and the like as used herein, when referring to a binding partner (e.g. protein, nucleic acid, antibody or other affinity capture agent, etc.), refers to a binding reaction of two or more binding partners with high affinity and/or complementarity to ensure selective hybridization under designated assay conditions. Typically, specific binding will be at least three times the standard deviation of the background signal. Thus, under designated conditions the binding partner binds to its particular "target" molecule and does not bind in a significant amount to other molecules present in the sample.

[0033] "Sequencing", "sequence determination" and the like means determination of information relating to the nucleotide base sequence of a nucleic acid. Such information may include the identification or determination of

partial as well as full sequence information of the nucleic acid. Sequence information may be determined "with varying degrees of statistical reliability or confidence. In one aspect, the term includes the determination of the identity and ordering of a plurality of contiguous nucleotides in a nucleic acid, "High throughput digital sequencing" or "next generation sequencing" means sequence determination using methods that determine many (typically thousands to billions) of nucleic acid sequences in an intrinsically parallel manner, i.e. where DNA templates are prepared for sequencing not one at a time, but in a bulk process, and where many sequences are read out preferably in parallel, or alternatively using an ultra-high throughput serial process that itself may be parallelized. Such methods include but are not limited to pyrosequencing (for example, as commercialized by 454 Life Sciences, Inc., Branford, Conn.); sequencing by ligation (for example, as commercialized in the SOLiD™ technology, Life Technology, Inc., Carlsbad, Calif.); sequencing by synthesis using modified nucleotides (such as commercialized in TruSeq™ and HiSeq™ technology by Illumina, Inc., San Diego, Calif., HeliScope™ by Helicos Biosciences Corporation, Cambridge, Mass., and PacBio RS by Pacific Biosciences of California, Inc., Menlo Park, Calif.), sequencing by ion detection technologies (Ion Torrent, Inc., South San Francisco, Calif.); sequencing of DNA nanoballs (Complete Genomics, Inc., Mountain View, Calif.); nanopore-based sequencing technologies (for example, as developed by Oxford Nanopore Technologies, LTD, Oxford, UK), and like highly parallelized sequencing methods.

[0034] The term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. Several equations for calculating the T_m of nucleic acids are well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation, T_m=81.5+0.41 (% G+C), when a nucleic acid is in aqueous solution at 1M NaCl (see e.g., Anderson and Young, Quantitative Filter Hybridization, in *Nucleic Acid Hybridization* (1985)). Other references (e.g., Allawi and SantaLucia, Jr., Biochemistry, 36:10581-94 (1997)) include alternative methods of computation which take structural and environmental, as well as sequence characteristics into account for the calculation of T_m.

DETAILED DESCRIPTION OF THE INVENTION

[0035] The practice of the techniques described herein may employ, unless otherwise indicated, conventional techniques and descriptions of organic chemistry, polymer technology, molecular biology (including recombinant techniques), cell biology, biochemistry, and sequencing technology, which are within the skill of those who practice in the art. Such conventional techniques include polymer array synthesis, hybridization and ligation of polynucleotides, and detection of hybridization using a label. Specific illustrations of suitable techniques can be had by reference to the examples herein. However, other equivalent conventional procedures can, of course, also be used. Such conventional techniques and descriptions can be found in standard laboratory manuals such as Green, et al., Eds., *Genome Analysis: A Laboratory Manual Series* (Vols. I-IV) (1999); Weiner, Gabriel, Stephens, Eds., *Genetic Variation: A Laboratory Manual* (2007); Dieffenbach, Dveksler, Eds., *PCR Primer: A Laboratory Manual* (2003); Bowtell and Sambrook, *DNA Microarrays: A Molecular Cloning Manual* (2003); Mount, *Bioinformatics: Sequence and Genome Analysis* (2004); Sambrook and Russell, *Condensed Protocols from Molecular Cloning: A Laboratory Manual* (2006); and Sambrook and Russell, *Molecular Cloning: A Laboratory Manual* (2002) (all from Cold Spring Harbor Laboratory Press); Stryer, *Biochemistry* (4th Ed.) (1995) W.H, Freeman, New York N.Y.; Gait, "*Oligonucleotide Synthesis: A Practical Approach*" (2002) IRL Press, London; Nelson and Cox, *Lehninger, Principles of Biochemistry* (2000) 3rd Ed., W. H. Freeman Pub., New York, N.Y.; and Berg, et al., *Biochemistry* (2002) 5th Ed., W. H. Freeman Pub., New York, N.Y., all of which are herein incorporated in their entirety by reference for all purposes.

[0036] Note that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a nucleic acid" refers to one or more nucleic acids, and reference to "the assay" includes reference to equivalent steps and methods known to those skilled in the art, and so forth.

[0037] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications mentioned herein are incorporated by reference for the purpose of describing and disclosing devices, formulations and methodologies that may be used in connection with the presently described invention.

[0038] Where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either both of those included limits are also included in the invention,

[0039] In the following description, numerous specific details are set forth to provide a more thorough understanding of the present invention. However, it will be apparent to one of skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features and procedures well known to those skilled in the art have not been described in order to avoid obscuring the invention.

The Invention in General

[0040] The assay systems of the invention provide spatially-encoded, multiplexed assays comprising 1) an assay capable of high levels of multiplexing with an efficient spatial encoding scheme; 2) instrumentation capable of delivering reagents according to a spatial pattern; and 3) decoding determined by a readout that is digital in nature. The assay systems of the invention detect the presence or absence and relative amount of a biological target or biological activity indicative of a biological target, as well as the location of the biological target or activity in a biological sample, e.g., a tissue section or other biological structure disposed upon a support such as a microscope slide or culture dish.

[0041] The assay system further provides instrumentation with an ability to deliver reagents in a spatially-defined pattern. This instrumentation, together "with software, reagents and protocols, provides a key component of the highly innovative assay system of the invention, allowing for measurement of numerous biological targets or activities in a meaningful spatial environment, including gene expression and peptide localization, An encoding scheme used in these assay systems allows one to determine the location of biological targets or activity (or lack thereof) in the biological samples after the products of the multiplexed assay are removed from the biological sample and pooled for analysis. Decoding of the encoding scheme can be performed by, e.g., next-generation sequencing, which easily provides millions to trillions of data points at low cost. The assay results such as the amount or activity of biological targets can then be mapped back to specific location in the biological sample. The assay systems open a new analytical window into the complex spatial patterns of cellular function and regulation in biological samples.

[0042] A simplified overview of the assay system **100** of the present invention is provided at FIG. **1**, At step **110**, a biological sample affixed to a support is provided. The biological sample contains biological targets of interest. Biological targets can include any molecule of interest, such as nucleic acids (including, e.g. RNA transcripts, genomic DNA sequences, cDNAs, amplicons, or other nucleic acid sequences) and proteins, enzymes and the like. At step **120**, encoded probes are delivered to the biological sample according to a known spatial pattern. Encoded probes comprise probes, which can interact "with biological targets of interest, and coding tags, which identify the positions in the sample of the biological targets being assayed, and thus can be used to link assay results back to locations in the sample. Coding tags in most embodiments are oligonucleotides. However, coding tags may also be mass tags, fluorescent labels, or other moieties.

[0043] In some embodiments, the probe and coding tag portions of the encoded probe are pre-coupled before being delivered to the biological sample, For example, in the case where the encoded probes are oligonucleotides, both the probe and coding tag sequence can be synthesized as a single oligonucleotide. Alternatively, the probe and coding tag portions of the encoding probes can be synthesized or obtained separately and combined before delivery to the biological sample (e.g., two separate oligonucleotides can be synthesized and coupled by, e.g., ligation; or an antibody and an oligonucleotide can be prepared separately and conjugated before delivery to the biological sample). Also, as is described in FIGS. **2-5**, the probes and the coding tags (in encoding oligonucleotides) are synthesized separately, and are delivered to the biological sample at different steps (e.g., probes first and coding tags thereafter, or vice versa) in the assay.

[0044] At step **130**, the encoded probes are allowed to react or interact with the biological targets, i.e., conditions are provided to allow e.g., oligonucleotides to hybridize to nucleic acid targets, enzymes to catalyze reactions with protein targets, antibodies to bind epitopes, etc. In the case where the biological targets are nucleic acids, the encoded probes are typically oligonucleotides and hybridize to the target nucleic acids. In the case that the biological targets are proteins, the encoded probes typically are aptamers, small molecules, or oligonucleotide-conjugated proteins that interact with target proteins by binding to them or by reacting with them (that is, one of the proteins is a substrate for the other). Encoding oligonucleotides may be coupled to the probes (proteins) by conjugation, chemical or photo-cross-linking via suitable groups and the like.

[0045] Once encoded probes interact with the biological targets, the encoded probes that interacted with the biological targets must be separated from the encoded probes that did not interact with the biological targets at step **140**. In the case where the biological targets are nucleic acids and the encoded probes are oligonucleotides, the separation can be accomplished by, e.g., washing the unhybridized encoded probes from the sample. Similarly, for other assays that are based on affinity binding, including those using aptamer, small molecule, and protein probes, washing steps can be used to remove low affinity binders. In the case where the probe is transformed via interaction with the target, e.g., in the case of a peptide, e.g., via cleavage by a protease or phosphorylation by a kinase, it is convenient to collect, all encoded probes--both encoded probes that interacted with the biological targets and were transformed and encoded probes that were not transformed. After collection or pooling, an antibody or other affinity capture agent can be used to capture probes that were transformed by addition of a moiety (e.g., a phosphate group). In cases where probes have been transformed via cleavage, the transformed probes can be separated, e.g., by capturing the non-transformed probes via a tag that is removed from the transformed probes during the transformation (e.g., by cleavage), or by adding a new tag at the site of cleavage.

[0046] Once the reacted (transformed) or interacted encoded probes are separated from the unreacted or uninteracted encoded probes, the sequence of the reacted and/or interacted encoded probes is determined by, preferably, sequencing. The sequence of the encoded probes allows the mapping of the assay results back to locations in the biological sample.

[0047] FIG. **2** provides a simplified overview of an assay system of the present invention embodying an efficient implementation of a combinatorial coding scheme for the encoding of spatial information. For purposes of this overview, the probes are oligonucleotides, but as explained elsewhere, other types of probes can also be used. In step **210**, a biological sample affixed to a support, e.g., a tissue sample or other biological structure, is provided. In step **220**, one or more oligonucleotide probes are delivered to the biological sample, where the oligonucleotide probes are capable of hybridizing with biological targets in the biological sample. In step **230**, the oligonucleotide probes are allowed to interact with (hybridize to) the nucleic acid targets; that is, appropriate conditions are provided where oligonucleotide probes can hybridize to the target nucleic acids.

[0048] In step **240**, the oligonucleotide probes that did not hybridize to target nucleic acids are removed, and thereby separated from oligonucleotide probes that did hybridize to target nucleic acids. In this embodiment, separation can be accomplished by, e.g., washing the sample to remove unhybridized oligonucleotide probes. Next, in step **250**, encoding oligonucleotides (the encoding agents) are delivered to the biological sample according to a chosen spatial pattern, where the encoding oligonucleotides comprise coding tags that are used to encode the location of biological targets in the biological sample. Note that in contrast to the assay

system of FIG. **1**, here the probes and encoding agents (encoding oligonucleotides) are delivered in separate steps. In step **260**, the encoding oligonucleotides are coupled to the oligonucleotide probes to create encoded probes. In this case where the probes are oligonucleotides, the encoding oligonucleotides may be coupled to the oligonucleotides probes by, e.g., ligation. Alternatively, the information in the encoding oligonucleotides can be transferred by using a DNA polymerase to extend a probe oligonucleotide that acts as a primer, and thereby copy and incorporate the sequence of the encoding oligonucleotides.

[0049] In step **270**, the sequence of the coding tags in the encoded probes as well as the sequence or a portion of the sequence of the probe itself is determined, and in step **280**, the target nucleic acids are mapped back to the biological sample. In some embodiments, the abundance of sequences reveals the relative quantity of biological targets at the location. Although this embodiment shows the individual steps in a particular order, so as to better explain the invention, the precise order of the steps can be varied. For example, steps **220** and **250** can be combined, so that a mixture of the probes and encoding oligonucleotides is delivered according to a chosen spatial pattern. Coupling step **260** can then be carried out immediately after the combined steps **220** and **250**, or concomitantly with them. In this case, step **240** would then occur after step **260**. It can therefore be appreciated that the two key results of this series of steps, i.e., the location-specific encoding of probe molecules and the separation of probe molecules based on their ability to interact with corresponding target molecules, can be accomplished with some flexibility in the implementation of the particular steps. Similarly, there is considerable flexibility in the design of the coding scheme. As described infra, the assays of the invention are particularly amenable to combinatorial methods.

[0050] Thus, the present invention provides an ability to look at many different biological targets in many locations, providing the resolution of in situ hybridization with the highly-parallel data analysis of sequencing. In some embodiments, the sum of the multiple biological targets being assayed and the multiple sites in the biological sample is greater than 20, in other embodiments, the sum of the multiple biological targets being assayed and the multiple sites in the biological sample is greater than 50, in other embodiments, the sum of the multiple biological targets being assayed and the multiple sites in the biological sample is greater than 100, greater than 500, 1,000, 10,000, 25,000, 100,000, 500,000, 1,000,000. It will be appreciated that, due to the spatial encoding dimension of the invention, even much larger numbers can be contemplated. For example, assaying 10,000 targets per location×10,000 locations would generate 10^{8} different assays, and even larger numbers than these can easily be contemplated, particularly if spatial locations with resolution on the order of that of single cells are utilized. Further, in embodiments where high-throughput digital sequencing is employed, the sequences of at least 1,000 encoding probes are typically determined in parallel. More typically, using a digital readout, it is desirable to obtain multiple sequence reads for each assay (defined by a probe and a spatial location code). It is desirable to obtain an average of at least 3 copies per assay, and more typically at least 10 or at least 30 copies per assay, depending on the design of the experiment and requirements of the assay. For a quantitative readout with suitable dynamic range, it may be desirable to obtain at least 1,000 reads per assay. Therefore, if 1,000,000 assays are carried out, the number of sequence reads may be 1 billion or more. With high-throughput digital sequencing, and allowing for redundancy, the sequence of at least 10,000 encoding probes are determined in parallel, or the sequence of at least 100,000, 500,000, 1,000,000, 10,000,000, 100,000,000, 1,000,000,000 or more encoding probes are determined in parallel.

Assays

[0051] The assay portion of the assay systems of the present invention comprise the following general steps: delivering probes and encoding agents where the encoding agents (in some embodiments pre-coupled to the probes) are delivered to the sample according to a known spatial pattern, allowing the probes to interact or react with biological targets in the sample, and, if the probes and encoding agents have not been pre-coupled, coupling the encoding agents to probes.

[0052] The samples of the present invention include virtually any biological sample or samples that can be affixed to a support or provided essentially in a two-dimensional manner, where the ability to tie an assayed biological target or activity back to the location within the biological sample is important. Exemplary biological samples include tissue sections (e.g., including whole animal sectioning and tissue biopsies), cell populations on slides or culture dishes, and the like. The assay systems of the invention are particularly advantageous in that they are compatible with numerous biological sample types, including fresh samples, such as primary tissue sections, and preserved samples including but not limited to frozen samples and paraformalin-fixed, paraffin-embedded (FFPE) samples. An important aspect of the assay systems of the invention is that the biological samples are immobilized on a substrate surface having discrete, independently measurable areas.

[0053] The biological targets to be detected can be any biological molecules including but not limited to proteins, nucleic acids, lipids, carbohydrates, ions, or multicomponent complexes containing any of the above. Examples of subcellular targets include organelles, e.g., mitochondria, Golgi apparatus, endoplasmic reticulum, chloroplasts, endocytic vesicles, exocytic vesicles, vacuoles, lysosomes, etc.

[0054] In some particular embodiments, the assay system is used to analyze nucleic acids, e.g., by genotyping, quantitation of DNA copy number or RNA transcripts, localization of particular transcripts within samples, and the like. FIG. **3** illustrates an overall scheme for an exemplary assay for, e.g., detecting single nucleotide polymorphisms (SNPs) that can be used with the assay system of the invention. In FIG. **3**, two oligonucleotide probes are provided. Each oligonucleotide probe comprises a target-specific region (located on either side of the SNP to be analyzed) seen at **305** and **307**, and ligation regions, seen at **301** and **303**, The oligonucleotide probes are allowed to hybridize to a target nucleic acid (not shown) in the biological sample. At step **302**, one of the oligonucleotide probes is extended to incorporate the SNP sequence and ligated to the other probe to form an extended probe comprising target nucleic acid region **309** and ligation regions **301** and **303**.

[0055] Two encoding agents, both comprising a coding tag (seen at **315** and **317**), a ligation region (seen at **311** and **313**), and a primer region (seen at **319** and **321**) are combined with and ligated to the extended probe at step **304**

to form an encoded target-specific oligonucleotide. Again, in contrast with FIG. **1**, the probes and encoding agents are delivered at separate steps. Doing so allows use of the combinatorial embodiments described infra. In preferred embodiments, the encoding oligonucleotides within a pair of encoding oligonucleotides ligate specifically to one side of the target sequence or the other (i.e., 5' or 3' of the target sequence). Also, typically, the ligation and primer regions of the encoding oligonucleotides and probes are universal; that is, the set of ligation and primer regions used in constructing the probes and encoding oligonucleotides are constant, and only the target-specific regions of the probes and the coding tags of the encoding oligonucleotides differ. However, again in alternative embodiments, the ligation and primer regions are not universal and differ between probes and encoding agents.

[0056] Following ligation, the encoded probes are eluted, pooled, and, optionally, sequencing adapters are added to the encoded probes via PCR. In alternative embodiments, sequencing primers may be ligated to the encoding oligonucleotides, or sequencing primer sequences can be included as part of the encoding oligonucleotide. As seen in FIG. **3**, each sequencing adapter comprises primer region **319** or **321**, compatible with the primer regions **319** and **321** on the encoded probes. The final construct comprising first adapter **327**, first primer region **319**, first coding tag **315**, ligation regions **311** and **301**, target region **309**, ligation regions **313** and **303**, second coding tag **317**, second primer region **325** and second adapter **329** is now ready for input into a digital high-throughput sequencing process.

[0057] A combination of extension and ligation reactions are exemplified in FIG. **3**, but it should be appreciated that a variety of reactions may be used to couple the encoding oligonucleotides to the target-specific oligonucleotides, including ligation only (e.g., for oligonucleotides that hybridize to contiguous portions of the target nucleic acid sequence). Alternatively, an assay utilizing an additional oligonucleotide, such as in the GOLDENGATE® assay (see Fan, et al., Cold Spring Symp. Quant. Biol., 68:69-78 (2003); (Ilumina, Inc., San Diego, Calif.)), may be employed.

[0058] In other embodiments, the assay system of the invention also can be used to analyze peptides or proteins, the presence of antibodies, enzymatic and other protein activities, posttranslational modifications, active and non-active forms of peptides, as well as peptide isoforms in a biological sample. Accordingly, the probes may comprise an active region of an enzyme, a binding domain of an immunoglobulin, defined domains of proteins, whole proteins, synthetic peptides, peptides with introduced mutations, aptamers and the like.

[0059] In certain aspects, the probes are substrates for enzymes or proenzymes, e.g., kinases, phosphatases, zymogens, proteases, or fragments thereof. In certain aspects, the probes are phosphorylation substrates used to detect proteins involved in one or more signal transduction pathways, e.g., a kinase or a phosphatase. In another specific aspect of the invention, the probes are specific protease substrates that associate only with individual proteases or classes of proteases. In other aspects, the probes are different processed forms, isoforms and/or domains of an enzyme. Protein-based probes are typically conjugated or otherwise linked to oligonucleotide encoding agents. The oligonucleotide encoding agents in this case would also include a nucleotide sequence component that allows for identification of the protein probe.

[0060] In certain aspects, the present invention provides assays for evaluating differences in the amount and/or activity of biological targets between different locations in a sample and/or between samples. The method includes determining a plurality of encoded results from the biological sample and evaluating the differences in quantity of the biological targets at each location in the biological sample.

Combinatorial Embodiments

[0061] To maximize the efficiency of encoding, a combinatorial approach using pairs of coding tags in the encoding oligonucleotides can be used. By de-coupling the target-specific information and the coding tags, the number of oligonucleotides required is dramatically reduced, with a concomitant decrease in cost.

[0062] FIG. **4** illustrates a general mechanism for one embodiment of a combinatorial encoding scheme of the assay systems of the invention, where nucleic acids in a representative tissue section (shown at **416**) are assayed. FIG. **4** at A shows two target-specific/encoding oligonucleotide constructs **420** and **422** (e.g., formed between steps **302** and **304** of FIG. **3**) specifically bound to a target nucleic acid **402** of interest, The first encoded probe **420** comprises coding tag **408**, associated with, e.g., a universal priming site for amplification of the assay products or an adapter to enable identification of the coding identifiers using sequencing technologies **404**. The second encoded probe **422** comprises coding tag **406**, associated with, e.g., a universal priming site for amplification of the assay products or an adapter to enable identification of the coding identifiers using sequencing technologies **410**.

[0063] FIG. **4** at B shows the spatial pattern that may be used for twenty different coding tags, a**1** through a**10** (coding tag **406** on encoded probe **420**) and b**1** through b**10** (coding tag **408** encoded probe **422**). Coding tag a**1**., for example, is deposited on the biological sample in ten discrete areas or spots (shown as the first horizontal line of spots in **412**). Coding tag a**2** is deposited on the biological sample in ten spots on the second horizontal line in **412**. Coding tag a**3** is deposited on the biological sample in ten spots on the third horizontal line in **412**, and so on. Whereas the "a" tags are deposited in ten horizontal rows, the "b" tags are deposited in ten vertical rows as shown in **414**. For example, coding tag b**1** is deposited on the biological sample in ten discrete spots in the first vertical row of **414**, coding tag b**2** is deposited on the biological sample in ten discrete spots in the second vertical row of **414**, and so on. Using such a configuration allows for twenty coding tags to uniquely define **100** different locations on the biological sample,

[0064] FIG. **4** at C shows a representative tissue section **416** coincident with coding tag grid **418**. The arrows show how the "a" coding tags and the "b" coding tags are deposited on grid **418** that is coincident with tissue section **416**. If, once sequenced, coding tags a**1** and b**4**, e.g., are associated with a target nucleic acid sequence, then that target nucleic acid sequence (i.e., biological target) was present in the tissue section at location a**1**, b**4**.

[0065] FIG. **5** provides a simplified, specific example of the encoding scheme of the assay systems of the invention. FIG. **5** shows encoding oligonucleotides **510**, comprising a**1**,

a**2**, a**3**, a**4** and b**1**, b**3**, b**3** and b**4**. Target-specific oligonucleotides (TSOs) (probes) **1** and **2** are shown at **520**. A deposit or dispensing scheme is shown at **530**. Like the grid exemplified in FIG. **4**, encoding oligonucleotides a**1** through a**4** are deposited in spots in a pattern (here, in a vertical pattern), and encoding oligonucleotides b**1** through b**4** are deposited in spots in a pattern (here, a horizontal pattern). The grid though shown as a square with spots is actually a deposition pattern on a biological sample (not shown) such as tissue section **416** shown in FIG. **4**.

[0066] The target-specific oligonucleotides are delivered to the biological sample, where the target-specific oligonucleotides hybridize to target nucleic acids in the biological sample if target nucleic acids are present. Unhybridized target-specific oligonucleotides are then removed, e.g., by washing. The encoding oligonucleotides are then delivered to the biological sample according to the spatial pattern shown at **530**. The encoding oligonucleotides are ligated (or, e.g., extended and ligated) to any target-specific oligonucleotides that hybridized to the target nucleic acid in the biological sample, the ligated constructs are then eluted from the biological sample, pooled, and sequencing adapters are added through, e.g., PCR or ligation, if the sequences were not previously included in the encoding oligonucleotides. The ligated constructs are sequenced by, e.g., high throughput or "next generation" sequencing.

[0067] The pool of resulting sequences is shown at **540**. A sequence readout was obtained for target-specific oligonucleotide **1** only at a**4**b**1**, a**4**b**2**, a**1**b**3**, a**2**b**3**, a**3**b**3**, a**4**b**3** and a**4**b**4** (positions shown with horizontal lines). A sequence readout was obtained for target-specific oligonucleotide **2** only at a**1**b**1** (position shown with vertical lines). A sequence readout was obtained for both target-specific oligonucleotides **1** and **2** at positions a**2**b**1**, a**3**b**1**, a**1**b**2**, a**2**b**2**, and a**3**b**2** (positions shown with cross-hatching). No sequence readout was obtained for either target-specific oligonucleotides at a**1**b**4**, a**2**b**4** or a**3**b**4** (positions shown without shading). Thus, in the biological sample on which the assay took place the first target nucleic acid was detected in a large portion of the left side and at the bottom of the biological sample, the second target nucleic acid was detected only in the upper left portion of the biological sample, and neither target nucleic acid was detected in the upper right portion of the biological sample. The differential expression of the two target nucleic acids now can be mapped back to the biological sample and to the biological structures or cell types in these locations in the biological sample.

[0068] In addition to location information, information relating to relative abundance of the encoded tags can be obtained. For example, if it is found that there are ten times as many a**4**T**1**b**1** sequences occurring in the data set as compared to a**4**T**1**b**2** sequences, this would indicate that target nucleic acid sequence **1** is ten times more abundant at the a**4**T**1**b**1** location than at the a**4**T**1**b**2** location.

[0069] In the case of nucleotide analysis as shown in FIG. **3**, by ligating the coding tags directly to target-specific oligonucleotides, only 2n target-specific oligonucleotides are needed for n targets. For example, using the combinatorial approach outlined in FIG. **2**, assaying 100 different targets at 10,000 spatial locations would require 2×100 target-specific oligonucleotides and 2×100 encoding oligonucleotides. The total count of assay oligonucleotides would be only 400 (200 target-specific and 200 encoding), not counting universal primers. In contrast, if the coding oligonucleotides were not decoupled from the target-specific oligonucleotides, (n×X positional codes)+(n×Y positional codes) would be needed, or in the above example, 20,000 oligonucleotides, not counting universal primer sequences. Moreover, though the embodiments shown in FIGS. **2**-**5** depict a combinatorial scheme using two encoding agents (coding tags), three, four or more encoding agents and coding tags may be used, and attached to the probe or one another by varying means and in varying combinations of steps.

[0070] Due to the spatial encoding aspect of the assay system of the invention, a large amount of information can be generated with even a modest number of assays. For example, five or more biological targets assayed at five or more positions in the sample generates 25 or more combinations. Using digital sequencing as a readout, the optimum number of sequence reads per combination depends on the sensitivity and dynamic range required, and can be adjusted, For example, if for each combination on average 100 reads are sampled, the total for 25 combination is 25,000 reads. If 1,000 targets are assayed at 1,000 locations with an average sampling depth of 1,000, then 10^9 reads are required. These numbers, although large, are within the capacity of intrinsically parallel digital sequencing methods, which can generate datasets of billions or even trillions of reads in a reasonable timeframe and at a very low cost per read. Therefore, by varying the numbers of positions interrogated or biological targets assayed, or both, and using digital sequencing, large amounts of information can be obtained. In specific aspects, multiple locations are interrogated for two or more biological molecules.

Reagent Delivery Systems

[0071] The reagent delivery system of the invention includes instrumentation that allows the delivery of reagents to discrete portions of the biological sample, maintaining the integrity of the spatial patterns of the encoding scheme. Reagent delivery systems of the assay systems of the invention comprise optional imaging means, reagent delivery hardware and control software. Reagent delivery can be achieved in a number of different ways. It should be noted that reagent delivery may be to many different biological samples at one time. A single tissue section has been exemplified herein; however, multiple biological samples may be affixed and analyzed simultaneously. For example, pions of a tissue sample can be analyzed in parallel and the data combined to build a 3D map.

[0072] Integral to the assay system of the invention is instrumentation that allows for spatial patterning of reagents onto the biological sample. Technologies for formulating and delivering both biological molecules (e.g. oligonucleotides or antibodies) and chemical reagents (e.g., small molecules or dNTPs) are known in the art, and uses of these instrument systems are known to one skilled in the art and easily adaptable to the assay systems of the invention. One example of a suitable reagent delivery system is the Labcyte™ Echo acoustic liquid handier, which can he used to deliver nanoliter scale droplets containing biological molecules with high precision and reproducibility. One skilled in the art could incorporate this reagent delivery device into the overall system, using software to specify the locations to which reagents should be delivered.

[0073] Other instruments that can be used for the deposition of agents and/or coding identifiers onto biological

samples include, but are not limited to, ink jet spotting; mechanical spotting by means of pin, pen or capillary; micro contact printing; photochemical or photolithographic methods; and the like. For several applications, it may be preferred to segment or sequester certain areas of the biological samples into one or more assay areas for different reagent distributions and/or biological target determination. The assay areas may be physically separated using barriers or channels.

[0074] In one exemplary aspect, the reagent delivery system may be a flow-based system. The flow-based systems for reagent delivery in the present invention can include instrumentation such as one or more pumps, valves, fluid reservoirs, channels, and/or reagent storage cells. Reagent delivery systems are configured to move fluid to contact a discrete section of the biological sample. Movement of the reagents can be driven by a pump disposed, for example, downstream of the fluid reagents. The pump can drive each fluid reagent to (and past) the reaction compartment. Alternatively, reagents may be driven through the fluid by gravity. US Pub. Nos. 20070166725 and 20050239192 disclose certain general-purpose fluidics tools that can be used with the assay systems of the invention, allowing for the precise manipulation of gases, liquids and solids to accomplish very complex analytical manipulations with relatively simple hardware.

[0075] In a more specific example, one or more flow-cells can be attached to the substrate-affixed biological sample from above. The flow-cell can include inlet and outlet tubes connected thereto and optionally an external pump is used to deliver reagents to the flow-cell and across the biological sample. The flow cells are configured to deliver reagents only to certain portions of the biological sample, restricting the amount and type of reagent delivered to any specific section of the biological sample.

[0076] In another aspect, a microfluidic system can be integrated into the substrate upon which the biological sample is disposed or externally attached on top of the substrate. Microfluidic passages for holding and carrying fluid may be formed on and/or above the planar substrate by a fluidics layer abutted to the substrate. Fluid reagents can be selected and delivered according to selective opening and closing of valves disposed between reagent reservoirs.

[0077] Pumps generally include any mechanism for moving fluid and/or reagents disposed in fluid. In some examples, the pump can be configured to move fluid and/or reagents through passages with small volumes (i.e., microfluidic structures). The pump can operate mechanically by exerting a positive or negative pressure on fluid and/or on a structure carrying fluid, electrically by appropriate application of an electric field(s), or both, among other means. Exemplary mechanical pumps may include syringe pumps, peristaltic pumps, rotary pumps, pressurized gas, pipettors, etc. Mechanical pumps may be micromachined, molded, etc. Exemplary electrical pumps may include electrodes and may operate by electrophoresis, electroendoosmosis, electrocapillarity, dielectrophoresis (including traveling wave forms thereof), and/or the like.

[0078] Valves generally include any mechanism for regulating the passage of fluid through a channel. Valves can include, for example, deformable members that can be selectively deformed to partially or completely close a channel, a movable projection that can be selectively extended into a channel to partially or completely block a channel, an electrocapillary structure, and/or the like.

[0079] An open gasket can be attached to the top of the biological sample and the sample and reagents can be injected into the gasket. Suitable gasket materials include, but are not limited to, neoprene, nitrile, and silicone rubber. Alternatively, a watertight reaction chamber may be formed by a gasket sandwiched between the biological sample on the substrate and a chemically inert, water resistant material such as, but not limited to, black-anodized aluminum, thermoplastics (e.g., polystyrene, polycarbonate, etc), glass, etc.

[0080] In an optional embodiment, the assay system comprises imaging means to determine features and organization of the biological sample of interest. The images obtained, e.g., may be used to design the deposition pattern of the reagents, Imaging means are optional, as an individual can instead view the biological sample using, e.g., a microscope, analyze the organization of the biological sample, and specify a spatial pattern for delivery assay reagents. If included, the delivery system can comprise a microcircuit arrangement including an imager, such as a CCD or IGFET-based (e.g., CMOS-based) imager and an ultrasonic sprayer for reagent delivery such as described in US Pub. No. 20090197326, which is incorporated herein by reference. Also, it should be noted that although FIGS. **4** and **5** illustrate using a x,y grid configuration, other configurations can be used, such as, e.g., following the topology of a tissue sample; targeting certain groups of cells, cell layers and/or cell types in a tissue, and the like.

[0081] In yet another alternative, the reagent delivery system controls the delivery of reagents to specific patterns on a biological sample surface using semiconductor techniques such as masking and spraying. Specific areas of a biological sample can be protected from exposure to reagents through use of a mask to protect specific areas from exposure. The reagents may be introduced to the biological sample using conventional techniques such as spraying or fluid flow. The use of masked delivery results in a patterned delivery scheme on the substrate surface.

[0082] In a preferred aspect of the invention, the reagent delivery instrumentation is based on inkjet printing technology. There are a variety of different ink-jetting mechanisms (e.g., thermal, piezoelectric) and compatibility has been shown with aqueous and organic ink formulations. Sets of independently actuated nozzles can be used to deliver multiple reagents at the same time, and very high resolutions are be achieved.

[0083] In order to target specific sites of interest, an informative image of the biological sample to be assayed may be used to assist in the reagent delivery methods and associated encoding scheme. Sample regions of the biological sample can be identified using image processing (e.g., images of cell types differentiated by immunohistochemistry or other staining chemistries) integrated with other features of the assay system. In some aspects, software is used to automatically translate image information into a reagent delivery pattern. A mechanism to register and align very precisely the biological sample for reagent delivery is thus an important component of the assay systems of the invention. Mechanisms such as the use of fiducial markers on slides and/or other very accurate physical positioning systems can be adapted to this purpose.

[0084] The invention preferably comprises a complete suite of software tailored to the assay system. Optionally,

oligonucleotide design software is used to design the encoding nucleotides (and in embodiments where nucleic acids are assayed, the target-specific oligonucleotides) for the specific assay to be run, and may be integrated as a part of the system. Also optionally, algorithms and software for reagent delivery and data analysis (i.e., sequence analysis) may be integrated to determine assay results. Integrated data analysis is particularly useful, as the type of dataset that is generated may be massive as a consequence of scale. Algorithms and software tools that are specifically designed for analysis of the spatially-associated data generated by the assay systems, including pattern-analysis software and visualization tools, enhance the value of the data generated by the assay systems.

[0085] In certain aspects, the assay system comprises processes for making and carrying out the quality control of reagents, e.g., the integrity and sequence fidelity of oligonucleotide pools. In particular, reagents are formulated according to factors such as volatility, stability at key temperatures, and chemical compatibility for compatibility with the reagent delivery instrumentation and may be analyzed by instrumentation integrated within the assay system.

Sequencing

[0086] Numerous methods can be used to identify the coding tags and probe sequences in the encoded probes of the assay systems of the invention. The coding tags can be detected using techniques such as mass spectroscopy (e.g., Maldi-T of, LC-MS/MS), nuclear magnetic resonance imaging, or, preferably, nucleic acid sequencing. Examples of techniques for decoding the coding tags of the present invention can be found, for example, in US Pub. No. 20080220434, which is incorporated herein by reference. For example, the coding tags may be oligonucleotide mass tags (OMTs or massTags). Such tags are described, e.g., in US Pub. No. 20090305237, which is incorporated by reference in its entirety. In yet another alternative, the encoded probes can be amplified and hybridized to a microarray. This would require separate amplification reactions to be carried out, in which each amplification is specific to a particular spatial code or subset of codes, accomplished by using code-specific primers. Each amplification would also incorporate a different resolvable label (e.g. fluorophor). Following hybridization, the relative amounts of a particular target mapping to different spatial locations in the sample can he determined by the relative abundances of the resolvable labels.

[0087] In one particularly preferred aspect, the resulting coding tags according to the assay system are substrates for high-throughput, next-generation sequencing, and highly parallel next-generation sequencing methods are used to confirm the sequence of the coding tags, for example, with SOLiD™ technology (Life Technologies, Inc.) or Genome Ananlyzer (Illumina, Inc.). Such next-generation sequencing methods can be carried out, for example, using a one pass sequencing method or using paired-end sequencing. Next generation sequencing methods include, but are not limited to, hybridization-based methods, such as disclosed in e.g., Drmanac, U.S. Pat. Nos. 6,864,052; 6,309,824; and 6,401,267; and Drmanac et al, U.S. patent publication 2005/0191656; sequencing-by-synthesis methods, e.g., U.S. Pat Nos. 6,210,891; 6,828,100; 6,969,488; 6,897,023; 6,833,246; 6,911,345; 6,787,308; 7,297,518; 7,462,449 and 7,501,245; US Publication Application Nos. 20110059436; 20040106110; 20030064398; and 20030022207; Ronaghi, et al, Science, 281: 363-365 (1998); and Li, et al, Proc. Natl. Acad. Sci., 100: 414-419 (2003); ligation-based methods, e.g., U.S. Pat. Nos. 5,912,148 and 6,130,073; and U.S. Pat. Appln Nos. 20100105052, 20070207482 and 20090018024; nanopore sequencing e.g., U.S. Pat. Appln Nos. 20070036511; 20080032301; 20080128627; 20090082212; and Soni and Meller, Clin Chem 53: 1996-2001 (2007)), as well as other methods, e.g., U.S. Pat. Appln Nos. 20110033854; 20090264299; 20090155781; and 20090005252; also, see, McKernan, et al., Genome Res., 19:1527-41 (2009) and Bentley, et al., Nature 456:53-59 (2008), all of which are incorporated herein in their entirety for all purposes.

Applications of Assay System

[0088] It will be apparent to one skilled in the art upon reading the present disclosure that there are numerous important areas of biological research, diagnostics, and drug development that will benefit from a high throughput multiplexed assay system that can measure simultaneously the amount and spatial location of a biological target in a biological sample. For example, combining the ability to estimate the relative abundance of different RNA transcripts with the ability to reconstruct an image of spatial patterns of abundance across many locations, which may be as small as or even smaller than individual cells, in a tissue enables many different areas of basic research. The following are exemplary uses and are by no means meant to be limiting in scope.

[0089] In one example, 3-dimensional patterns of gene expression are determined by analyzing a series of tissue sections, in a manner analogous to image reconstruction in CT scanning, Such a method can be used to measure changes in gene expression in disease pathology, e.g., in cancerous tissue and/or a tissue upon injury, inflammation or infection. With the assay systems of the invention, more detailed information on gene expression and protein localization in complex tissues is obtained, leading to new insights into the function and regulation both in normal and diseased states, and provides new hypotheses that can be tested. For example, an assay system of the invention may enable some of the insights gained from many individual studies and larger programs like ENCODE (Birney, et al., Nature, 447: 799-816 (2007)) and modENCODE to be integrated at the tissue level. The assay systems also aid computational efforts to model interacting networks of gene expression in the field of systems biology.

[0090] The assay systems also provide a novel approach to analysis of somatic variation, e.g., somatic mutations in cancer or variability in response to infectious organisms. For example, tumors are typically highly heterogeneous, containing cancer cells as well as genetically normal cells in an abnormal local environment. Cancer cells undergo mutation and selection, and in this process it is not unusual for local clones to develop. Identifying relatively rare somatic mutations in the context of tumors may enable the study of the role of key mutations in the selection of clonal variants. Transcriptional patterns associated with angiogenesis, inflammation, or other cancer-related processes in both cancer and genetically normal cells can be analyzed for insights into cancer biology and assist in the development of new therapeutic agents for the treatment of cancers. In another example, individuals have varying susceptibility to

infectious organisms, and the assay systems of the invention can be used to study the interaction between microbes and tissues or the various cell types within the tissue.

[0091] Importantly, in addition to providing spatially-associated information, the invention allows a great increase in the sensitivity of detecting rare mutations, as signal to noise can he dramatically increased since only a small location is assayed in any given reaction. In a typical assay for rare mutations in a mixed sample, the sample is treated in bulk, i.e., nucleic acids are extracted from many cells into a single pool. Thus, if a mutation is present in one cell in 10,000, it must be detected against a background of normal DNA from ~10,000 cells. In contrast, with the assay systems of the invention many cells can be analyzed, but individual cells or small groups of cells would be identified by the spatial coding system. Therefore, in the assay systems of the present invention, background is reduced by orders of magnitude, greatly increasing sensitivity. Furthermore, the spatial organization of mutant cells can be observed, which may be particularly important in detecting key mutations in tissue sections in cancer. Already molecular histological analyses are yielding insights into cancer biology and may have potential for use in diagnostics. The technology of the invention promises to greatly increase the power of such approaches.

[0092] The present invention provides assays, assay systems, and methods of using such assays in spatially encoded biological assays. The invention provides an assay system comprising one or more agents provided in defined spatial patterns on a substrate surface, and a detection system for identifying the presence or absence, relative amount, and location of a biological molecule. Such biological molecules include, but are not limited to, nucleic acids, peptides, carbohydrates, cellular components, and the like. The assay system is a novel multiplexing approach, as it allows multiple molecules and their respective multiple locations to be identified in a single system using a unique encoding scheme. This encoding scheme uses both molecule-specific binding agents and coding identifiers to provide a practical and cost-effective determination of information on multiple biological molecules, including specific positional information of such molecules in a biological sample, e.g., a tissue section. The single molecule detection analysis using the encoding system also allows relative amounts of biological molecules to be detected, thus providing information on expression levels, sequestering in specific locales, and the like.

[0093] The assay systems detect the presence or absence, and relative amount, of a biological molecule at more than one spatial location in a sample. In addition, the assays provide methods for doing this for multiple biological molecules simultaneously. The assay systems utilize one or more binding agents that specifically bind to the biological molecule of interest and unique coding identifiers associated with specific binding agents. The detection system utilizes a method for identifying the presence and spatial address of the agent binding based on the positive and/or negative results that are obtained using detection of the agent and identifier and the encoding scheme of the spatial patterns on the substrate surface. In a specific aspect, the encoding scheme employs limited reagent delivery to the spatial patterns on the substrate surface, and access of the coding identifiers and/or binding agents to portions of the sample is controlled through such limited delivery.

[0094] In one aspect, the assay system detects the presence or absence and spatial location of a biological molecule based on the positive and/or negative results that are obtained using limited reagent delivery and the encoding scheme of the spatial patterns on the substrate surface.

[0095] The assay system and methods of the invention are based on relational, solid-state substrates with positions that represent specific spatial locations within a biological sample, e.g., a cell, organelle or tissue. The ability to use encoding features to represent locations allows high-throughput analysis of the presence or absence, and relative amount, of a biological molecule at more than one spatial location in a sample. The encoding features also allow provide assaying of multiple biological molecules at these multiple locations simultaneously.

[0096] A primary feature of the invention is the preservation of the spatial organization of elements in a sample of interest through the use of an encoding scheme. For example, the assay may be designed to preserve the relative position of cells in a tissue, and the assay may interrogate the individual cells for genomic DNA variation (including epigenetic modifications), and RNA and protein expression.

[0097] In one specific aspect, the encoding scheme of the assay system comprises the use of two or more coding patterns, each comprising regions defined by spatial patterns on the substrate surface. For example, the assay system can utilize an encoding scheme that comprises a 2-dimensional grid format based on the discrete positioning of the binding agents in the substrate surfaces. In another example, the spatial patterns may be based on more randomized cell locations, e.g., the patterns on the substrate surface follow an underlying biological structure rather than a strict, x,y grid pattern. This aspect includes systems with two or more substantially identical spatial patterns using different binding agents and/or coding identifiers, as well as systems having different patterns for different agents and/or coding identifiers. The encoding scheme of the systems can be controlled by delivery of different reagents to discrete regions on the substrate surfaces, which allows different reactions to take place on substantially similar agents of known location on the substrate surfaces.

[0098] In one specific aspect, the invention provides high resolution, high-throughput analysis of nucleic acids and/or expression levels that provides both detection and spatial identification of large numbers of nucleic acids, e.g., DNA or RNA.

[0099] In another specific aspect, the invention provides high resolution, high-throughput analysis of proteins that provides both detection and spatial identification of large numbers of such proteins, e.g., kinases or proteases.

[0100] Numerous reagent delivery systems can be used with the assay system of the invention. The primary criteria of such reagent delivery systems is the ability to direct delivery of specific agents based on spatial patterns on the substrate surface.

[0101] In one preferred aspect, the encoding scheme utilizes a reagent delivery system based on printing and informatics technologies to implement the spatial patterns used for identification and localization of the biological materials. For example, the patterns found in the encoding scheme may be created using ink jet printing technology to provide reagents at specific locations on one or more substrate surfaces. The desired patterns are set out in specific coding patterns on the substrate surface.

[0102] In certain aspects of the invention, the binding agents are immobilized directly to the substrate surface, and the location of the binding agents is known or determined prior to use of the substrate surface in the assay system. In another aspect, the binding agents are immobilized onto beads or other separate structural elements that are then provided in known locations on the substrate surface. In yet another aspect, the binding agents may be provided in or on features of the substrate surface, e.g., provided in wells or channels.

[0103] In specific aspects of the invention, the binding agents are nucleic acids immobilized directly or indirectly to the substrate surface, e.g., directly through the use of amino groups on the substrate surface or indirectly through the use of a linker. The location of the nucleic acid sequences is known or determined prior to use of the substrate surface in the assay system. In another specific aspect, the nucleic acids may be immobilized directly or indirectly onto beads that are then provided in known locations on the substrate surface. In yet another aspect, the nucleic acids may be provided in or on features of the substrate surface, e.g., provided in wells.

[0104] In these aspects involving nucleic acid agents, any methods of sequence determination can be used, e.g., sequencing, hybridization and the like. In a preferred aspect, nucleic acid sequencing, and preferably next-generation sequencing, is used to decode the spatial encoding scheme in the assay system of the invention. This provides a very wide dynamic range for very large numbers of assays, allowing for efficient multiplexing.

[0105] In some aspects, the assay utilizes two or more oligonucleotides, the oligonucleotides comprising a universal primer region and a region that correlates specifically to a single spatial pattern within the spatial encoding scheme. In a specific aspect, the assay comprises two allele specific oligonucleotides and one locus specific oligonucleotides. These oligonucleotides allow the identification of specific SNPs, indels or mutations within an allele. This is useful in the identification of genetic changes in somatic cells, genotyping of tissues, and the like.

[0106] In other specific aspects of the invention, the binding agents are peptides. In one aspect, these peptides are associated directly or indirectly to known locations on a substrate surface, e.g., using binding protein pairs or through oligonucleotide linkers complementary to oligonucleotides on the substrate surface. In another aspect, the binding agents are peptides are immobilized directly or indirectly onto beads or other separate structural elements that are then provided in known locations on the substrate surface. In yet another aspect, the peptides may be provided in or on features of the substrate surface, e.g., provided in wells.

[0107] In yet other specific aspects of the invention, the binding agents are chemical entities (e.g., small molecules) that are coded, e.g. using sequence tags or mass spectroscopy tags as coding identifiers. In one aspect, these chemical entities can be are immobilized directly to the substrate surface. In another aspect, the binding agents are immobilized onto beads or other separate structural elements that are then provided in known locations on the substrate surface. In yet another aspect, the binding agents may be provided in or on features of the substrate surface, e.g., provided in wells.

[0108] The assay system of the invention can utilize various detection mechanisms, based on the molecules to be detected and the reagents needed for such detection system. Exemplary methods that can be used with the assay systems of the invention are described in more detail below.

The Invention in General

[0109] The assay system and methods of the invention are based on relational methods that allow extraction of data to detect the presence or absence and relative amount of a biological molecule, and the location of this molecule in a sample having a distinct structure, e.g., a tissue section or other biological structure with distinct locations of specific biological molecules. The encoding scheme used in these systems corresponds to the structural elements of the sample, and the information obtained using a two-dimensional coding system is indicative of the spatial addresses of these molecules in a sample of interest.

[0110] Integral to the assay system of the invention is a method for spatial patterning of reagents. Technologies for formulating and delivering both biological molecules (e.g. DNA or antibodies) and chemical reagents (e.g., small molecules or dNTPs) have already been demonstrated, and use of these systems will be available to one skilled in the art and easily adaptable upon reading this specification.

[0111] The assay design of the invention provides an accurate and easily scalable spatial encoding system. The ability to deliver reagents in a spatially defined pattern together with software, reagents and protocols comprises a novel and highly innovative assay system for spatial analysis of various biological molecules and activities. This allows the assays to measure numerous biological functions in a meaningful spatial environment, including functions such as gene expression and peptide localization. The systems provide the potential to open a new analytical window into the complex spatial patterns of cellular function and regulation in biological systems.

[0112] The biological molecules to be detected can be any biological molecules such as proteins, nucleic acids, lipids, carbohydrates, ions, or multicomponent complexes containing any of the above. Further examples of subcellular objects include organelles, e.g., mitochondria, Golgi apparatus, endoplasmic reticulum, chloroplast, endocytic vesicle, exocytic vesicles, vacuole, lysosome, etc.

[0113] FIG. **1** illustrates such a target-specific assay system for identification of nucleic acid sequences in a sample. In this system **101**, two reagents **120**, **122** that specifically bind to a biological molecule of interest are associated with coding identifiers **106**, **108** that encode for a spatial location in the sample. These coding identifiers **106**, **108** are optionally associated with sites that assist in their identification in the assay format, e.g., universal priming sites **104**, **110** for amplification of the assay products or adapters to enable identification of the coding identifiers and the binding agents using sequencing technologies. The sample that is tested, here shown as a tissue section **116** is encoded using the combination of the patterns **112**, **114** created using the separate coding identifiers **106**, **108** which provide a two dimensional code **118** that shows the location of any positive detection of the biological molecule **102** as well as quantifying the biological molecule **102** at each location assayed in the tissue.

[0114] The assay systems of the invention are particularly advantageous in that they are compatible with numerous samples types, such as fresh samples, such as primary tissue sections, and preserved samples including but not limited to

frozen samples and paraformalin-fixed, paraffin-embedded (FFPE) samples. An important aspect of the assay systems of the invention is that the binding agents are immobilized on a substrate surface in discrete, independently measureable areas. These discrete areas can be formed by spatially selective deposition of the binding agents on the substrate surface. Numerous methods can be used for the deposition of the agent and the coding identifiers associates with the agent. For example, the coding identifiers can be delivered together or separately from the agent. If delivered together they can be attached (e.g., synthesized as a single molecule or attached through ligation or a chemical coupling mechanism) or simply mixed together to be attached after delivery to the substrate. In a preferred aspect, the agent and the coding identifier are made separately, mixed together for attachment, and delivered either attached or as a mixture to be attached on the surface. In a specific aspect the binding agents are delivered generally over the substrate surface and the coding identifiers are delivered in a pattern-specific manner.

[0115] Examples of methods that can be used for deposition of agents and/or coding identifiers onto the substrate surface include, but are not limited to, ink jet spotting, mechanical spotting by means of pin, pen or capillary, micro contact printing, fluidically contacting the measurement areas with the biological or biochemical or synthetic recognition elements upon their supply in parallel or crossed micro channels, upon exposure to pressure differences or to electric or electromagnetic potentials, and photochemical or photolithographic immobilization methods.

[0116] For several applications, it may be preferred to arrange the substrates into segments of one or more measurement areas for reagent distribution and agent determination. These regions may be physically separated using barriers or channels. They may still comprise several additional discrete measurement areas with agents that are different or in different combination from each other.

[0117] In certain aspects, the present invention provides a method, e.g., a machine-based method, for evaluating changes in the presence and/or location of a biological molecule over time. The method includes providing a plurality of encoded array results representative of the biological molecule over time and evaluating the differences in detection and/or localization of the biological molecules.

Nucleic Acid Detection and Localization

[0118] In a particular aspect, the assay system is used to analyze nucleic acids, e.g genotyping, gene expression analysis, localization of particular transcripts within samples, and the like.

[0119] Genotyping may be performed using any technique known to those of skill in the art. Preferred techniques permit rapid, accurate determination of multiple variations with a minimum of sample handling. Some examples of suitable techniques involve but are not limited to direct DNA sequencing, capillary electrophoresis, hybridization, allele-specific probes or primers, single-strand conformation polymorphism analysis, nucleic acid arrays, bead arrays, restriction fragment length polymorphism analysis, cleavage fragment length polymorphism analysis, random amplified polymorphic DNA, ligase detection reaction, heteroduplex or fragment analysis, differential sequencing with mass spectrometry, atomic force microscopy, pyrosequencing, FRET (e.g., TaqMan (Applied Biosystems, Inc., Foster City, Calif.) and Molecular Beacon (Stratagene, La Jolla, Calif.) assays), and other related techniques. Several methods for DNA sequencing are well known and generally available in the art. See, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, New York) (2001); Ausubel, et al., Current Protocols in Molecular Biology (John Wiley and Sons, New York) (1997), Twyman, et al. (2003) "Techniques Patents for SNP Genotyping", Pharmacogenomics 4(1):67-79; and Kristensen, et al. (2001) "High-Throughput Methods for Detection of Genetic Variation", BioTechniques 30(2):318-332. For details on the use of nucleic acid arrays (DNA chips) for the detection of, for example, SNPs, see U.S. Pat. No. 6,300,063 issued to Lipshultz, et al., and U.S. Pat. No. 5,837,832 to Chee, et al., HuSNP Mapping Assay, reagent kit and user manual, Affymetrix Part No. 90094 (Affymetrix, Santa Clara, Calif.). The molecular inversion probe (MIP) assay format (Hardenbol et al., 2003) is another example of a highly multiplexable assay that may be used with the assay systems of the invention.

[0120] In one exemplary and preferred method for analyzing nucleic acids using the assay system of the invention, the detection of nucleic acids uses two allele-specific oligonucleotides and a locus specific oligonucleotide. The assay methods are carried out according to the strategy outlined in FIG. **2** using next-generation sequencing or another highly parallel nucleic acid assay technology. In this assay, a set of two oligonucleotides is designed to hybridize to each target sequence, with a common oligonucleotide and two unique coding identifiers. The allele can be determined, e.g. by primer extension of the locus specific oligonucleotide. Following primer extension and ligation, an amplifiable template is formed with universal primer sequences at either end. Assay oligonucleotides are annealed to a template and enzymatic reactions are used to join the two oligonucleotides only when both are correctly annealed. The detection techniques and read out parameters used in this system of the invention include a much shorter tag than the oligonucleotides used in the assays that are based on capture by hybridization. These shorter tags are designed to be read out by sequencing or, preferably, used to ligate codes onto both ends of the fragment as illustrated in FIG. **2**.

[0121] In FIG. **2**, two target-specific assay oligonucleotides are ligated together **202** following in situ hybridization to target sequences. At the same time, encoding oligonucleotides containing tag sequence sets X and Y are ligated **204** to the target specific oligonucleotides. Oligonucleotides containing X ligate specifically to one side of the targeting construct and oligonucleotides containing Y ligate to the other. The oligonucleotides contain universal primer sites P**1** and P**2**. Following ligation, the constructs are eluted and, optionally, sequencing adapters can be attached **206**, e.g., via PCR.

[0122] In one preferred aspect, the final construct created from the assay method is a substrate for next-generation sequencing, and highly parallel next-generation sequencing methods are used to confirm the sequence of constructs. Such sequencing methods can be carried out, for example, using a one pass sequencing method or using paired-end sequencing. Next generation sequencing methods include, but are not limited to, hybridization-based methods, such as disclosed in Drmanac, U.S. Pat. Nos. 6,864,052; 6,309,824; and 6,401,267; and Drmanac et al, U.S. patent publication 2005/0191656, and sequencing by synthesis methods, e.g.,

Nyren et al, U.S. Pat. No. 6,210,891; Ronaghi, U.S. Pat. No. 6,828,100; Ronaghi et al (1998), Science, 281: 363-365; Balasubramanian, U.S. Pat. No. 6,833,246; Quake, U.S. Pat. No. 6,911,345; Li et al, Proc. Natl. Acad. Sci., 100: 414-419 (2003); Smith et al, PCT publication WO 2006/074351; use of reversible extension terminators, e.g., Turner, U.S. Pat. No. 6,833,246 and Turner, U.S. Pat. No. 6,833,246 and ligation-based methods, e.g., Shendure et al (2005), Science, 309: 1728-1739, Macevicz, U.S. Pat. No. 6,306,597; which references are incorporated by reference. Soddart et al., PNAS USA. 2009 Apr. 20; Xiao et al., Nat Methods. 2009 March; 6(3):199-201. Epub 2009 Feb. 8.

[0123] To maximize the efficiency of encoding, a combinatorial approach using pairs of oligonucleotides can be used. For example, with only two sets of 100 codes, a substrate can theoretically encode up to 10,000 locations. The number of assay oligonucleotides required is dramatically reduced, the cost decreased, and the robustness of the approach increased by decoupling the coding sequences from the genome-specific sequences. Alternative assay formats can also be used (e.g. ligation or primer extension followed by ligation).

[0124] By ligating the codes on separately, only 2n target-specific assay oligonucleotides are needed for n targets. For example, assaying 100 different targets at 10,000 spatial locations would require 2×100 targeting oligonucleotides and 2×100 encoding oligonucleotides, using a combinatorial approach outlined in FIG. **2**. The total count of assay oligonucleotides would be only 400 (200 target-specific and 200 encoding), not counting universal primers. In contrast, if the coding oligonucleotides were not decoupled, (n×X positional codes)+(n×Y positional codes) would be needed, or in the above example, 20,000 oligonucleotides, not counting universal primer sequences.

[0125] Due to the matrix system of the invention, a large amount of information can be obtained even using five or more positions interrogated for five or more biological molecules. By varying one or the other of these, large amounts of information can be obtained, both in terms of locations and/or specific biological In specific aspects, multiple locations are interrogated for two or more biological molecules. As an example, for each datapoint 1,000 reads may be sampled, for a total of −10E9 reads for 10E6 datapoints.

Peptide Detection Systems

[0126] The assay system of the invention can be used to analyze biological molecules using peptide agents that are associated with the substrate surface in a spatial pattern. Such peptides may comprise an active region of an enzyme, a binding domain of an immunoglobulin, defined domains of proteins, whole proteins, synthetic peptides, peptides with introduced mutations, etc.

[0127] The assay system of the invention allows the identification and spatial location of various forms of peptides, including isoforms and peptides that have undergone posttranslational modification. Importantly, certain aspects of the invention allow the identification of active versus non-active forms of such peptides in a sample. This allows the identification of the presence or absence of specific peptide isoforms, and also acts as a proxy for identification of peptide activity in a sample.

[0128] In certain aspects of the invention, the binding agents associated with the substrate surfaces of the assay system include substrates for enzymes or proenzymes, e.g., a kinase, a phosphatase, a zymogen, a protease, or a fragment thereof. In certain aspects, the binding agents associated with the substrate surfaces are phosphorylation substrates used to detect proteins involved with one or more signal transduction pathways, e.g., a kinase or a phosphatase. In another specific aspect of the invention, the binding agents are specific protease substrates that associate only with individual or classes of proteases. In other aspects, the binding agents on the substrate surface are different processed forms, isoforms and/or domains of an enzyme.

Reagent Delivery

[0129] The reagent delivery system of the invention can be any system that allows the delivery of reagents to discrete portions of the array in order to keep the integrity of the defined spatial patterns of the encoding scheme. Such discrete delivery can be achieved in a number of different ways.

[0130] In one exemplary aspect, the reagent delivery system can be a flow-based system. The flow-based systems for reagent delivery in the present invention can include one or more pumps, valves, fluid reservoirs, channels, and/or reagent storage cells. Such a reagent delivery system is configured to move fluid in contact with a discrete section of the substrate surface. Movement of the reagents can be driven through a fluid by a pump disposed, for example, downstream of the fluid reagents. The pump can drive each fluid reagent to (and past) the reaction compartment. Alternatively, the reagents may be driven through the fluid by gravity.

[0131] US Appln Nos. 20070166725 and 20050239192 disclose certain general-purpose fluidics tools that can be used with the assay systems of the invention. These allow the precise manipulation of gases, liquids and solids to accomplish very complex analytical manipulations with relatively simple hardware.

[0132] In a more specific example, one or more flow-cells can be attached to the substrate from above. The flow-cell can include inlet and outlet tubes connected thereto and optionally an external pump can be used to deliver the sample or reagents to the flow-cell and across the substrate. The flow cell is configured to deliver reagents only to certain portions of the array, restricting the amount and type of reagent delivered to any specific section of the array.

[0133] In another aspect, a microfluidic system can be integrated into the substrate or externally attached on top of the substrate. Microfluidic passages for holding and carrying fluid can be formed on and/or above the planar substrate by a fluidics layer abutted to the substrate. Fluid reagents can be selected according to selective opening and closing of valves disposed between reagent reservoirs.

[0134] Pumps generally include any mechanism for moving fluid and/or reagents disposed in fluid. In some examples, the pump can be configured to move fluid and/or reagents through passages with small volumes (i.e., microfluidic structures). The pump can operate mechanically by exerting a positive or negative pressure on fluid and/or on a structure carrying fluid, electrically by appropriate application of an electric field(s), or both, among others. Exemplary mechanical pumps may include syringe pumps, peristaltic pumps, rotary pumps, pressurized gas, pipettors, etc. The mechanical pumps may be micromachined, molded, etc. Exemplary electrical pumps can include electrodes and may

operate by electrophoresis, electroendoosmosis, electrocapillarity, dielectrophoresis (including traveling wave forms thereof), and/or the like.

[0135] Valves generally include any mechanism for regulating the passage of fluid through a channel. The valves can include, for example, deformable members that can be selectively deformed to partially or completely close a channel, a movable projection that can be selectively extended into the channel to partially or completely block the channel, an electrocapillary structure, and/or the like.

[0136] In yet another aspect, an open gasket can be attached to the top of the substrate and the sample and reagents can be injected into the gasket. Suitable gasket materials include, but are not limited to, neoprene, nitrile, and silicone rubber. Alternatively, a watertight reaction chamber formed by a gasket sandwiched between the substrate and a chemically inert, water resistant material such as, but not limited to, black-anodized aluminum, thermoplastics (e.g., polystyrene, polycarbonate, etc), glass, etc.

[0137] In a specific aspect of the present invention, the delivery system can compose a microcircuit arrangement including an imager, such as a CCD or IGFET-based (e.g., CMOS-based) imager and an ultrasonic sprayer for reagent delivery such as described in US Appln No. 20090197326, which is incorporated herein by reference.

[0138] In yet another aspect of the invention, the reagent delivery system controls the delivery of reagents to specific patterns on a substrate surface using semiconductor techniques such as masking and spraying. Specific areas of a substrate surface can be protected from exposure to reagents through use of a mask to protect specific areas from exposure. The reagents may be introduced to the substrate using conventional techniques such as spraying or fluid flow. The use of the masked substrate delivery results in a patterned delivery scheme on the substrate surface.

[0139] In a preferred aspect of the invention, the reagent delivery instrumentation is based on inkjet printing technology. There are a variety of different ink-jetting mechanisms (e.g., thermal, piezoelectric) and compatibility has been shown with aqueous and organic ink formulations. Sets of independently actuated nozzles can be used to deliver multiple reagents at the same time, and very high resolutions can be achieved.

Software for Use in the Assay System

[0140] In order to target specific sites of interest, an informative image of the biological section to be analyzed can be used to assist in the reagent delivery methods and associated encoding scheme. Sample regions can be identified using image processing (e.g., images of cell types differentiated by immunohistochemistry or other staining chemistries) integrated with the other features of the assay system. In some aspects, software is used to automatically translate this information into a reagent delivery pattern. A mechanism to register and align very precisely the biological sample in a targeting system is thus a preferred component of the assay systems of the invention. Mechanisms such as the use of fiducial markers on slides and other very accurate physical positioning systems can be adapted to this purpose.

[0141] Additional software components will also be key components that will be part of a complete assay system. The invention thus preferably comprises a complete suite of software tailored to the assay system. Optionally, oligonucleotide design software will be customized for the specific assay to be run, and may be integrated as a part of the system. Also optionally, algorithms and software for data analysis may be integrated to assist in determination of results of the assays. This can be especially useful, as the type of dataset that will be generated will be novel, particularly as a consequence of scale. The ability to provide algorithms and software tools that are specifically designed for analysis of spatially-associated data for significant patterns, including pattern-analysis software and visualization tools, is a novel feature that will enhance the value of the data generated by the assay systems.

[0142] In certain aspects, the assay system will comprise processes for making and carrying out quality control of reagents, e.g., the integrity and sequence fidelity of oligonucleotide pools. In particular, reagents will need to be formulated for compatibility with the reagent delivery instrumentation. Factors such as volatility, stability at key temperatures, and chemical compatibility can be optimized by those skilled in the art upon reading the present disclosure.

Applications of Assay System

[0143] It will be apparent to one skilled in the art upon reading the present disclosure that there are numerous very important areas of biological research, diagnostics, and drug development that will benefit from a high throughput means to simultaneously measure the presence or absence and spatial location of a biological molecule in a sample. For example, this technology combining the ability to analyze semi-quantitatively the expression of many different genes with the ability to image the spatial organization of expression across many cells in a tissue is enabling for many different areas of basic research. The following are exemplary uses and are by no means meant to be limiting in scope.

[0144] In one example, 3-dimensional patterns of expression can be determined by analyzing a series of tissue sections, in a manner analogous to image reconstruction in CT scanning. This can be used to measure changes in gene expression in disease pathology, e.g., in cancerous tissue and/or a tissue upon injury, inflammation or infection. With the assay systems of the invention, more detailed information on gene expression and protein localization in complex tissues can be obtained. This may lead to new insights into the function and regulation both in normal and diseased states, and is likely to provide new hypotheses that can be tested. For example, a system of the invention may enable some of the insights gained from many individual studies and larger programs like ENCODE (Birney et al., 2007) and modENCODE to be integrated at the tissue level. The assay systems will also aid in computational efforts to model interacting networks of gene expression in the field of systems biology.

[0145] The assay systems also provide a novel approach that enables the analysis of somatic variation, e.g., somatic mutations in cancer or variability in response to infectious organisms. For example, tumors are typically highly heterogeneous, containing cancer cells as well as genetically normal cells in an abnormal local environment. Cancer cells undergo mutation and selection, and in this process it is not unusual for local clones to develop. Identifying relatively rare somatic mutations in the context of tumors may enable the study of the role of key mutations in the selection of clonal variants. Transcriptional patterns associated with

angiogenesis, inflammation, or other cancer related processes in both cancer and genetically normal cells can be analyzed for insights into cancer biology and assist in the development of new therapeutic agents for the treatment of cancers.

[0146] In another example, different people have varying susceptibility to infectious organisms, and much of this may be to underlying genetic differences in individuals and/or populations. Identifying these differences will aid in an understanding of the underlying disease pathologies and assist in the development of vaccines or therapeutics to prevent or ameliorate these disease states.

[0147] Importantly, in addition to providing spatially associated information, the technology of the invention will allow a great increase in the sensitivity of detecting rare mutations. The reason is that signal to noise can be dramatically increased because the approach of the invention assays a small location in any given reaction. In a typical assay for rare mutations in a mixed sample, the sample is treated in bulk, i.e. nucleic acids are extracted from many cells into a single pool. Thus, if a mutation is present in 1 cell in 10,000, it must be detected against a background of normal DNA from ~10,000 cells. In contrast, with the systems of the invention many cells can be analyzed, but individual cells or small groups of cells would be identified by the spatial coding system. Therefore, the background can be reduced by orders of magnitude, greatly increasing sensitivity. Furthermore, the spatial organization of mutant cells can be observed. This may be particularly important in detecting key mutations in tissue sections in cancer. Already, molecular histological analyses are yielding insights into cancer biology and may have potential for use in diagnostics (Choe et al., 2003). The technology of the invention promises to greatly increase the power of such approaches.

EXAMPLES

[0148] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventor regards as his invention, nor are they intended to represent or imply that the experiments below are all of or the only experiments performed. It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

[0149] Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees centigrade, and pressure is at or near atmospheric.

Example 1

Initial Proof of Concept of Encoding Scheme

[0150] As an initial proof of concept, a model system is developed using a microarray to demonstrate a working single-plex assay. The basic design validates the concept of the assay, and establishes a working assay prior to addressing issues related to the analysis of a more complicated biological sample. Conventional sequencing is used as a readout for this proof of concept.

[0151] A microarray is used as a proxy for a tissue section. The target sequences of the microarray are fully specified, so that the composition of the targets are known and can be varied systematically. Synthetic oligonucleotide templates are attached to a glass slide via a 5' amino modification. Each slide has a single oligonucleotide template sequence, and the assays that are carried out may employ either ligation, or extension followed by ligation as this may be useful in determining certain polymorphisms.

[0152] Once the in situ part of the assay is complete, the reaction products are eluted and analyzed by qPCR to determined presence or absence of a product and estimate yield, and by conventional sequencing to determine the structure of the assay products. The single plex assays that are tested include appropriate positive and negative controls, and a single nucleotide variant (SNV) to check ability to discriminate single base variants.

Example 3

Scalability

[0153] The complexity of the assay system is increased to establish scalability of the assay for use in high throughput studies. Scalability of both the spatial encoding and assay systems is demonstrated by carrying out a 24-plex×24-site assay using a microarray model system.

[0154] The amount of biological target, here a DNA target sequence, at each assay location is systematically varied on microarray substrate. For example, in a microarray with 50 micron spot size (center to center), a 1 mm^2 area contains ~400 spots. The region around each site is optionally occupied by a region that is devoid of these spots to allow individual resolvability of the target sequences. Alternatively, the spots may be clustered, with two or more directly adjacent spots surrounded by or adjacent to a region that is devoid of target sequences.

[0155] In order to demonstrate that spatial encoding is accurate, the sites comprise different target compositions to show that the assay readout matches the expected composition of each site. With 24 target sequences, a simple digital pattern is made with each site having a different set of 12 targets present and 12 targets absent, to make a binary code (0=absent, 1=present). The assay readout is then determined to show that the detected regions match the expected signal after spatial decoding. In this particular example, the code space is large enough (2^{24}) so that even a few errors would not result in different codes being mixed up. Moreover, this design allows identification of errors and allows an estimation not only of accuracy of spatial encoding but also of accuracy calling the presence or absence of target sequences. In an exemplary aspect, a 4×4 arrangement of 16 sequences is used for the array configuration. A white square indicates that the sequence is absent and a black square that it is present, i.e. 8 of the 16 possible sequences are present in this sample. In a different sample, a different pattern of absent and present sequences can be constructed. In this way, unique patterns are associated with spatial locations so that the accuracy of spatial encoding can be measured.

[0156] The ability to detect quantitative differences is evaluated by generating dose-response curves for each of the

24 assays that are carried out at each site in a 24-site assay. This allows estimation of the limit of detection, dynamic range, and power to detect a given fold-change across the range.

[0157] In one aspect, a latin square design is used to represent individual targets at different ratios by varying the number of features for each target. In other words, with multiple spots in a site, the number of spots allocated to each of the 24 target sequences can be varied and each of the 24 sites can have a different composition. A 1×3 inch microarray is sufficiently large to permit multiple replicates. This larger set of 24 sequences will require deconvolution, and this is accomplished using high throughput techniques such as next-generation sequencing technologies (e.g., SOLiD™ technology (Life Technologies, Inc., Carlsbad, Calif.) or Genome Analyzer (Illumina, Inc., San Diego, Calif.)). The use of the 24-plex assay demonstrates both the accuracy of spatial encoding and decoding, and the quantitative response of the assay system.

Example 3

Adaptation of the Assay to Preserved Samples

[0158] Genomic DNA is assayed as a proof of concept for assaying RNA, as it provides a way to establish a single-copy reference signal. Once a working assay is developed for FFPE samples, it is adapted to an RNA assay. To this end, assay oligonucleotide concentrations are assayed to ensure compatibility with high multiplexing. Assuming a cell diameter of 10 microns, and delivery of a 10 micron diameter reagent droplet to an individual cell, the volume of the droplet will be ~500 μl and can contain ~3×10^{11} molecules at a 1 μM concentration. To assay 1,000 target sequences in 10,000 cells, ~2,000 targeting oligonucleotides would be required in a droplet. Therefore, each droplet could contain ~160 million copies of each assay oligo, a vast excess over the few thousand target sequences in a cell.

[0159] The handling of small absolute numbers of product molecules generated from very small or compromised samples are enhanced to counter the issue of low recovery efficiency; that is, elution is efficient and losses resulting from adsorption of molecules to surfaces are prevented. An approach to addressing the latter issue is to include a carrier material, such as glycogen or carrier nucleic acids.

Example 4

Adapting the Assay to a Biological Sample

[0160] A control RNA template is immobilized to a solid support in order to create an artificial system. The assay is performed using T4 DNA ligase, which can repair nicks in DNA/RNA hybrids. Assays are carried out on matched slides, or different sections of the same slide, where in one case gDNA is assayed and in the other RNA is assayed. When assaying gDNA the slide can be pretreated with RNase, and when assaying RNA the slide is pretreated with DNase. Results of the assay are confirmed by extracting gDNA or RNA and quantitating the relative amounts by qPCR or RT-qPCR respectively.

[0161] In order make the tissue section RNA assays as informative as possible, pre-existing information on expression levels in specific tissues to target transcripts across a range of abundances are used in the assay design. Both high abundance transcripts, as well as some medium and low abundance transcripts, are targeted to enable an initial assessment of the quantitative performance characteristics of the assay.

[0162] The preceding merely illustrates the principles of the invention. It will be appreciated that those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the invention and are included within its spirit and scope. Furthermore, all examples and conditional language recited herein are principally intended to aid the reader in understanding the principles of the invention and the concepts contributed by the inventors to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions. Moreover, all statements herein reciting principles, aspects, and embodiments of the invention as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents and equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure. The scope of the present invention, therefore, is not intended to be limited to the exemplary embodiments shown and described herein. Rather, the scope and spirit of present invention is embodied by the appended claims. In the claims that follow, unless the term "means" is used, none of the features or elements recited therein should be construed as means-plus-function limitations pursuant to 35 U.S.C. § 112, ¶6.

1. (canceled)

2. A method for determining a spatial location of a biological molecule of a tissue sample comprising

(a) providing a plurality of beads in a spatial pattern, wherein the plurality of beads comprise a plurality of binding agents, wherein a binding agent of the plurality of binding agents (i) comprises a coding identifier having a nucleic acid sequence and (ii) is configured to interact with a biological molecule of a tissue sample, wherein the coding identifier corresponds to a location in the spatial pattern at which the binding agent interacts with a biological molecule of a tissue sample;

(b) contacting the plurality of beads with the tissue sample such that the binding agent interacts with the biological molecule;

(c) using the coding identifier to identify the location in the spatial pattern where the binding agent interacted with the biological molecule, thereby determining a spatial location of the biological molecule of the tissue sample.

3. The method of claim **2**, wherein the plurality of beads is provided on a substrate surface.

4. The method of claim **3**, wherein the plurality of beads is provided in wells on the substrate surface.

5. The method of claim **2**, wherein the biological molecule is a nucleic acid.

6. The method of claim **5**, wherein the binding agent of the plurality of binding agents comprises a nucleic acid.

7. The method of claim **6**, wherein the binding agent interacts with the biological molecule via hybridization.

8. The method of claim **2**, wherein the binding agent and coding identifier are provided as a single molecule.

9. The method of claim **8**, wherein the binding agent and coding identifier are provided as a single molecule through nucleic acid synthesis.

10. The method of claim **8**, wherein the binding agent and coding identifier are provided as a ligation product.

11. The method of claim **8**, wherein the binding agent and coding identifier are provided as a single molecule through chemical coupling.

12. The method of claim **2**, wherein the plurality of binding agents comprises an additional binding agent, which (i) comprises an additional coding identifier and (ii) is configured to interact with another biological molecule of the tissue sample, wherein the additional coding identifier is different from the coding identifier of step (a)(i).

13. The method of claim **12**, wherein the additional coding identifier corresponds to a location in the spatial pattern at which the additional binding agent of the plurality of binding agents interacts with said another biological molecule of the tissue sample.

14. The method of claim **13**, wherein the location in the spatial pattern at which the binding agent interacts with the biological molecule and the additional binding agent interacts with said another biological molecule is different.

15. The method of claim **14**, wherein the spatial location of the biological molecule and said another biological molecule of the tissue sample is different.

16. The method of claim **2**, wherein step (c) comprises determining the spatial locations of a plurality of biological molecules of the tissue sample simultaneously.

17. The method of claim **2**, wherein step (c) comprises determining presence or absence of the biological molecule at the location at which it is identified.

18. The method of claim **2**, wherein step (c) comprises determining the relative amount of the biological molecule at the location at which it is identified.

19. The method of claim **2**, wherein step (c) comprises determining all or a portion of the nucleic acid sequence of the coding identifier.

20. The method of claim **2**, wherein step (c) comprises sequencing all or a portion of the nucleic acid sequence of the coding identifier.

21. The method of claim **2**, wherein spatial organization of cells in the tissue sample is preserved.

22. The method of claim **2**, wherein the biological molecule of the biological sample is selected from a protein, a nucleic acid, a lipid, a carbohydrate, and an ion.

23. The method of claim **22**, wherein the biological molecule is part of a multicomponent complex.

24. The method of claim **2**, wherein the tissue sample is a fresh tissue sample.

25. The method of claim **2**, wherein the tissue sample is a preserved tissue sample.

26. The method of claim **25**, wherein the preserved tissue sample is a frozen tissue sample.

27. The method of claim **26**, wherein the preserved tissue sample is a paraformalin-fixed paraffin-embedded (FFPE) tissue sample.

28. The method of claim **3**, further comprising analyzing gene expression in the tissue sample.

29. The method of claim **2**, wherein the spatial pattern is a defined spatial pattern.

30. The method of claim **3**, wherein the location of the binding agent is known prior to step (b).

31. The method of claim **3**, wherein the binding agent is immobilized directly or indirectly onto a bead of the plurality of beads.

* * * * *